27

ENCICLOPEDIA IBEROAMERICANA DE FILOSOFÍA

Filosofía de la lógica

Edición de
Raúl Orayen y Alberto Moretti

Editorial Trotta

Consejo Superior de Investigaciones Científicas



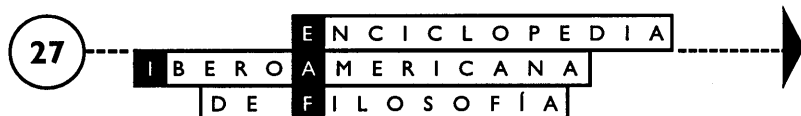


Filosofía de la lógica

Filosofía de la lógica

Edición de Raúl Orayen
y Alberto Moretti

Editorial Trotta
Consejo Superior de Investigaciones Científicas





Ferraz, 55. 28008 Madrid
Teléfono: 91 543 03 61
Fax: 91 543 14 88
E-mail: editorial@trotta.es
http://www.trotta.es

Departamento de Publicaciones
Vitruvio, 8. 28006 Madrid
Teléfono: 91 561 62 51
Fax: 91 561 48 51
E-mail: publ@orgc.csic.es

Diseño
Joaquín Gallego

ISBN: 84-87699-48-0 (Obra completa)
ISBN TROTTA: 84-8164-673-3 (vol. 27)
ISBN CSIC: 84-00-08288-5
NIPO: 653-04-016-3
Depósito Legal: M-52.217-2004

Impresión
Marfa Impresión, S.L.

La Enciclopedia IberoAmericana de Filosofía es un proyecto de investigación y edición, puesto en marcha por el Instituto de Filosofía del Consejo Superior de Investigaciones Científicas (Madrid), el Instituto de Investigaciones Filosóficas de la Universidad Autónoma de México y del Centro de Investigaciones Filosóficas (Buenos Aires), y realizado por filósofos que tienen el español por instrumento lingüístico.

Existe una pujante y emprendedora comunidad filosófica hispanoparlante que carece, sin embargo, de una obra común que orqueste su plural riqueza y contribuya a su desarrollo. No se pretende aquí una enciclopedia de filosofía española sino articular la contribución de la comunidad hispanoparlante a la filosofía, sea mediante el desarrollo cualificado de temas filosóficos universales, sea desentrañando la modalidad de la recepción a esos temas filosóficos en nuestro ámbito lingüístico.

La voluntad del equipo responsable de integrar a todas las comunidades filosóficas de nuestra área lingüística, buscando no sólo la interdisciplinariedad sino también la internacionalidad en el tratamiento de los temas, nos ha llevado a un modelo específico de obra colectiva. No se trata de un diccionario de conceptos filosóficos ni de una enciclopedia ordenada alfabéticamente sino de una enciclopedia de temas monográficos selectos. La monografía temática permite un estudio diversificado, como diverso es el mundo de los filósofos que escriben en español.

Queremos dejar constancia del agradecimiento debido a quienes formaron parte del Comité Académico, haciendo posible este proyecto, y que han fallecido: Carlos Alchourrón, Ezequiel de Olaso, J. L. López Aranguren, Fernando Salmerón y Javier Sasso.

La Enciclopedia IberoAmericana de Filosofía es el resultado editorial de un Proyecto de Investigación financiado por la Comisión Interministerial de Ciencia y Tecnología y por la Dirección General de Investigación Científica y Técnica del Ministerio de Educación y Ciencia. Cuenta también con la ayuda de la Consejería de Educación y Cultura de la Comunidad de Madrid.

CONTENIDO



PRESENTACIÓN

Alberto Moretti y Raúl Orayen

De quien se ocupa de caracterizar principios y criterios para clasificar los razonamientos como correctos o incorrectos, contribuyendo de este modo a entender nuestra práctica de producirlos y evaluarlos, es razonable decir que se ocupa de lógica. Como cualquier ocupación, ésta también genera el tipo de reflexión que solemos llamar filosófica. Una vez iniciada, es normal que esa reflexión no cese y conduzca a problemas y terrenos muy alejados del punto de partida. Si permanecemos en los estadios iniciales de una indagación filosófica motivada por la lógica, estaremos involucrados en los temas comúnmente considerados típicos de la filosofía de la lógica. Pero no es posible ni deseable que todos ellos estén explícitamente tratados cuando, como en este caso, sólo pretendemos ofrecer un panorama representativo de tipos de problemas y orientaciones característicos de la filosofía de la lógica. Las páginas que siguen contienen, pues, una selección de temas y enfoques básicos de este campo filosófico, que son o han sido tratados por buena parte de quienes lo cultivan en nuestro medio.

La práctica inferencial, tal como la identificamos paradigmáticamente, parece suponer, además de ciertas condiciones histórico-sociales, algún lenguaje natural y, desde luego, algún proceso psíquico. Desde fines del siglo diecinueve, la construcción de sistemas de lógica encargados de clarificarla requiere el diseño de estructuras sintácticas denominadas usualmente «lenguajes formales». Pero en otros tiempos estaba muy ligada a ciertas ideas sobre el funcionamiento de nuestra mente. El artículo de Agustín Rayo, «Formalización y lenguaje ordinario», plantea el problema del vínculo entre nuestros razonamientos, producidos en algún lenguaje social e históricamente constituido, y los recursos técnicos arbitrados para su examen, que

están asociados con algún *lenguaje formal*. Por su parte, Gladys Palau, en su contribución «Lógica y psicología», desarrolla el otro aspecto señalado, trazando el recorrido histórico de los intentos por fundar la lógica en los resultados de la *investigación psicológica*, mostrando también de qué modo perduran en ciertos ámbitos de la lógica contemporánea.

Nuestra comprensión general de la práctica de argumentar y de los razonamientos que produce (y la práctica incluye aceptar y rechazar razonamientos) se beneficia por el examen de las características específicas que esa actividad y sus productos presentan, en los diversos ámbitos en que se manifiestan. Uno de los más representativos lo ofrecen los discursos de carácter *aseverativo* donde pretendemos decir verdades acerca de algo. Pero, suponiendo que a veces logremos referirnos a algo en particular, ¿cómo es que lo logramos?, y ¿qué necesitamos creer sobre la idea de verdad para comprender las inferencias en esta región discursiva? La primera cuestión, la de la *referencia*, parece muy ligada a la de identidad: si vamos a estar en condiciones de hablar de más de una entidad (y la lógica cuantificacional proporciona una manera de hacerlo) deberemos ser capaces de distinguirlas, y si pensamos en esta capacidad requeriremos las ideas de diferencia y de identidad. El artículo de Rodolfo Gaeta, «Identidad y referencia», se dedica a plantear las dificultades de la relación entre esas nociones, estudiando algunos intentos de solución y mostrando su nexo con el concepto de necesidad. El artículo siguiente, «El concepto tarskiano de verdad», considera la segunda cuestión, la de la idea de *verdad* a la que apelamos cuando estudiamos cierto tipo de inferencias, exponiendo la manera en que es habitualmente caracterizada en los sistemas de lógica y sus vínculos con el análisis tradicional de raigambre aristotélica.

El motivo más inmediato para el desarrollo de la lógica es la preocupación por elaborar conceptos y criterios teóricos que permitan justificar la admisión de ciertas inferencias y el rechazo de otras, en consonancia con nuestras más arraigadas prácticas de evaluación. Hay inferencias de muy diversos tipos; lo atestiguan frases más o menos infelices como «lógica inductiva», «lógica deductiva», «razonamiento plausible», «abducción», «inferencia revocable». Para los casos más sencillos, aquellos en que pretendemos que la conclusión es inevitable en presencia de las premisas dadas, la tarea teórica se resume en la búsqueda de un concepto adecuado de *consecuencia lógica*. Mario Gómez Torrente desarrolla este problema en su artículo «La noción de consecuencia lógica». Describe las caracterizaciones de esa relación formuladas en términos de derivabilidad y en términos de verdad,

plantea el problema de su adecuación y destaca los vínculos de este concepto con los de forma lógica y expresión lógica.

En la construcción de teorías o sistemas de lógica se habla de, se definen, y quizás se descubran, proposiciones, oraciones, objetos y operaciones sintácticas, conjuntos, funciones, relaciones y estructuras. En el análisis de los juicios de corrección o incorrección referidos a razonamientos expresados en lenguajes naturales, se apela a la conveniencia o necesidad de hacer abstracción de ciertas peculiaridades con el propósito de destacar otras que parecen más importantes a esos fines. En este tipo de discusiones acerca de procesos y objetos, juega un importante papel la distinción entre entidades concretas y abstractas. Ignacio Angelelli, en «Abstracción», estudia el vuelco que la idea de proceso de *abstracción* sufrió durante el siglo veinte en relación, precisamente, con el desarrollo de la lógica, mostrando dos modos básicos de entenderlo: uno que reivindica la tradición y otro que no lo hace, ambos deudores de los métodos de la investigación lógica contemporánea. En cualquiera de los dos enfoques de la abstracción cabe la pregunta por la naturaleza de los *entes abstractos* presupuestos, postulados, construidos o descubiertos. El artículo de Eduardo Barrio, «Entidades abstractas», describe las líneas generales de la discusión actual en torno al problema de la existencia y naturaleza de estas entidades, refiriéndolo a los aparentes compromisos abstractos de la lógica, y señalando sus conexiones con la versión actual de la disputa sobre el realismo.

El momento culminante en la génesis de la lógica actual fue protagonizado por Frege y por el intento de representar la aritmética en términos puramente lógicos. La lógica que Frege usó para ese objetivo requiere un lenguaje con mayor poder expresivo que el supuesto por los sistemas de lógica elemental. Requiere extender el lenguaje elemental hasta permitir algo análogo a la cuantificación sobre conjuntos o propiedades de objetos del dominio y, de esta manera, estar en posición de formular principios lógicos adecuados para caracterizar la aritmética como lógica. Pero los principios que Frege propuso resultaron inconsistentes, y la responsabilidad mayor le cupo a una tesis acerca de entidades parecidas a los conjuntos de los que hoy hablamos. Como se ve, las extensiones de la lógica que involucran la posibilidad de referirnos a entidades típicamente abstractas, requieren mucha atención para no caer en antinomias. La contribución de Francisco Rodríguez Consuegra, «Tipos lógicos, lenguaje y filosofía», presenta la célebre solución russelliana a este problema: la *teoría de tipos*. En su exposición el autor destaca las restricciones adicionales que Russell reclamaba, concernientes al significado de las expresiones

lingüísticas y a la estructura de las entidades, y señala las dificultades derivadas para el proyecto logicista así como para su aplicación general al discurso cognoscitivo (dentro del cual está la propia formulación de la teoría de tipos). Un lenguaje alternativo al de esta teoría, y que permite representar las afirmaciones aritméticas y, en general, matemáticas, es el lenguaje conjuntista. Este lenguaje también parece permitirnos representar nuestras principales afirmaciones empíricas. Y, por otra parte, se ha difundido como lenguaje ortodoxo para dotar de una semántica rigurosa a los lenguajes lógicos elementales o de primer orden. Dada esta enorme capacidad expresiva, si ahora adoptamos axiomas especiales que caractericen a los conjuntos, ¿seguimos en el ámbito de la lógica?, ¿son los conjuntos objetos lógicos?, ¿es el de conjunto un concepto tan propio de la lógica como, digamos, el de negación? En «Lógica y teoría de conjuntos», Jesús Mosterín brinda un panorama de la riqueza conceptual de las *teorías de conjuntos* y muestra la estrecha relación entre lo que naturalmente llamaríamos lógicas, aunque de segundo orden, y las teorías axiomáticas de conjuntos, cuyas motivaciones fueron naturalmente matemáticas, vinculadas con el intento de evitar las paradojas originadas en las concepciones ingenuas de los conjuntos (aunque no por eso resultan meros trucos formales sin apoyo conceptual). Llama la atención, asimismo, sobre la incidencia de la teoría de conjuntos en la demostración habitual de las propiedades básicas de la lógica de primer orden, y en la determinación de las verdades lógicas elementales. Semejante vínculo reaviva la inquietud por clarificar qué sean estas entidades, abstractas, que llamamos conjuntos. Éste es el tema de la contribución de Ignacio Jané, «¿De qué trata la teoría de conjuntos?». Luego de exponer la *concepción iterativa* de los conjuntos, destacando los aspectos que la alejan de ser una construcción arbitraria, y planteándola como justificación de la axiomatización conocida como ZF, se ocupa de las interpretaciones realistas y no realistas de los conjuntos y bosqueja los detalles de una variante del segundo tipo.

Así como se han visto, o creído ver, motivos suficientes para ampliar la expresividad de los lenguajes formales y aumentar el número de principios lógicos, también se han visto, o creído ver, motivos para modificar algunos rasgos de los sistemas lógicos elementales y, consecuentemente, discutir la aplicabilidad general de la idea de consecuencia lógica que esos sistemas materializan. En el tomo de esta *Enciclopedia* dedicado a la lógica hay varios ejemplos de esta vertiente de la investigación. En el presente volumen, Raymundo Morado expone, en «Problemas filosóficos de la lógica no monotónica», las consideraciones que se han aducido para justificar la tesis de que cierta

relación no monótona entre premisas y conclusiones es una relación de consecuencia lógica, y que es de la mayor importancia para construir nuestra idea de racionalidad argumentativa, mereciendo entonces el esfuerzo por construir sistemas formales para caracterizarla. Desde otra perspectiva metodológica, el artículo de María Manzano, «Divergencia y rivalidad entre lógicas», estudia la posibilidad de que la lógica clásica (en particular su extensión multivariada) pueda resultar teóricamente adecuada, en algún sentido, en cualquier área, incluso en aquellas donde otros la creen impotente. El asunto naturalmente apunta a clarificar la idea misma *de sistema de lógica* y sugiere la tarea de comparar sistemas diversos y, por tanto, suscita la cuestión de la existencia de un instrumento lógico mediante el cual sea posible representar los diferentes sistemas cuyas propiedades estructurales se estudien.

Los compiladores agradecemos su trabajo a los autores y autoras, y su buena voluntad para colaborar en un proyecto colectivo que les impuso algunas restricciones y tareas inesperadas. A algunos, además, agradecemos su paciencia extraordinaria.

Queremos señalar, por último, que el primer diseño de contenidos y autores para el volumen se debe a Carlos Alchourrón, quien había asumido la tarea de compilarlo. De hecho, las primeras versiones de los artículos segundo, tercero y cuarto ya estaban en su poder cuando se produjo, en 1996, su lamentable e inesperado fallecimiento. Luego de variadas complicaciones hemos logrado completar su tarea, y nos gustaría que el trabajo finalizado pudiera verse, también, como un homenaje a su memoria.

FORMALIZACIÓN Y LENGUAJE ORDINARIO

Agustín Rayo

El lenguaje ordinario no siempre es apropiado para llevar a cabo nuestros objetivos. Por ejemplo, nuestros objetivos podrían requerir que nuestro discurso carezca de ambigüedades, y el lenguaje ordinario frecuentemente es ambiguo. Cuando *formalizamos* un enunciado lo parafraseamos utilizando un lenguaje formal —un cierto tipo de lenguaje artificial— más apropiado para el desarrollo de nuestros objetivos.

1. LENGUAJES FORMALES

Un lenguaje formal consiste de dos partes. La primera es un conjunto de símbolos básicos. La segunda es un conjunto de reglas, llamadas *reglas de formación*, que especifican qué secuencias de símbolos básicos han de contar como *fórmulas*.

No cualquier conjunto de símbolos básicos y cualquier conjunto de reglas de formación constituyen un lenguaje formal. Los símbolos básicos y las reglas de formación deben estar al alcance de seres finitos como nosotros, al menos en principio. Para esto no es necesario que los conjuntos sean finitos. Lo que requerimos es que sean *efectivamente decidibles*[1]. Dado un símbolo cualquiera, debe existir un algoritmo que nos permita determinar si el símbolo ha de ser contado entre nuestros símbolo básicos y, dada una regla cualquiera, debe ha-

1. En general, se dice que un conjunto es efectivamente decidible si existe un algoritmo que permita determinar, para cada objeto en el dominio de discurso, si ese objeto pertenece al conjunto.

ber un algoritmo que nos permita determinar si la regla ha de ser contada entre nuestras reglas de formación.

Como ejemplos de lenguajes formales, consideremos los llamados *cálculos proposicionales*. Los símbolos básicos de un cálculo proposicional son: (*a*) una serie (finita o infinita) de letras proposicionales: 'A_1', 'A_2', etc.; (*b*) conectivas proposicionales básicas: '$\wedge$' y '$\neg$'[2]; y (*c*) paréntesis: '(' y ')'. Las reglas de formación de un cálculo proposicional son las siguientes:

- Cualquier letra proposicional es una fórmula.
- Si ⌜ϕ⌝ y ⌜ψ⌝ son fórmulas, entonces también ⌜$(\phi \wedge \psi)$⌝ es una fórmula[3].
- Si ⌜ϕ⌝ es una fórmula, entonces también ⌜$(\neg\phi)$⌝ es una fórmula.
- Ninguna otra secuencia de símbolos básicos es una fórmula.

Así, cuando 'A_1', 'A_2' y 'A_3' son letras proposicionales, se sigue de nuestras definiciones que 'A_3', '$(\neg A_2)$', '$(A_1 \wedge A_3)$' y '$(A_1 \wedge (A_2 \wedge (\neg A_1)))$' son fórmulas, pero que '$A_3 \neg$' y '$A_1 \neg \wedge A_2$' no lo son.

En algunos casos se le asignan significados a las fórmulas de un lenguaje formal. En este caso se dice que el lenguaje formal está *interpretado*. Frecuentemente, la manera más fácil de asignarle significados a las fórmulas de un lenguaje formal es proporcionando una *traducción* del lenguaje formal a un lenguaje previamente significativo —un lenguaje natural, por ejemplo—. Así, podríamos traducir 'A_1' al castellano como 'la nieve es blanca', 'A_2' como 'el pasto es verde', '$(\neg A_1)$' como 'no es el caso que la nieve es blanca', y '$(A_1 \wedge A_2)$' como 'la nieve es blanca y el pasto es verde'.

Una traducción no es la única manera de asignarle significados a las fórmulas de un lenguaje formal. Otra manera de hacerlo es gene-

2. Además de conectivas proposicionales básicas, los cálculos proposicionales suelen incluir conectivas proposicionales derivadas, definidas a partir de las básicas. Por ejemplo, '$(A \vee B)$' suele entenderse como una abreviación de '$(\neg(\neg(A) \wedge \neg(B)))$', '$(A \supset B)$' suele entenderse como una abreviación de '$(\neg(A) \vee B)$', y '$(A \equiv B)$' suele entenderse como una abreviación de '$((A \supset B) \wedge (B \supset A))$'. En versiones distintas del cálculo proposicional, otras conectivas son consideradas como básicas y '$\wedge$' o '$\neg$' (o ambas) se consideran como derivadas.

3. Las esquinas, '⌜' y '⌝', fueron introducidas por primera vez en Quine, 1951. Para entender el uso de las esquinas basta con notar que cada instancia del siguiente esquema es verdadera cuando se escriben expresiones cualesquiera en el lugar de '__' y '...':

si ϕ es una expresión sin especificar, ⌜__ ϕ ...⌝ abrevia al resultado de escribir ϕ en el lugar de 'ϕ' en '__ ϕ ...'.

rando una *semántica formal*. Ése es el tema de la sección 5.1. En la sección 5.2 se discute la asignación de significados a un lenguaje formal por medio de axiomas, o reglas axiomáticas.

Los lenguajes formales suelen estar asociados con *sistemas deductivos*. Un sistema deductivo para un lenguaje formal F es un conjunto de reglas de la forma 'Si está permitido derivar cada fórmula en el conjunto Φ (tal vez vacío) a partir de las fórmulas en el conjunto P, entonces está permitido derivar la fórmula ψ a partir de las fórmulas en el conjunto Q'. Una *prueba finita* [infinita] en un lenguaje formal F con respecto a un sistema deductivo D es una secuencia finita [infinita] cada miembro de la cual es un par ordenado $<\phi, P>$ con las siguientes características: (*a*) ϕ es una fórmula de F, (*b*) P es un conjunto de fórmulas de F, y (*c*) alguna regla en D permite derivar ϕ a partir de P, bajo el supuesto de que para cada miembro anterior de la secuencia, $<\psi, Q>$, está permitido derivar ψ a partir de Q. Una fórmula ϕ de F es una *consecuencia (sintáctica)* a partir de P con respecto a D si $<\phi, P>$ aparece en alguna prueba en F con respecto a D. Una fórmula ϕ de F es un *teorema* de F con respecto a D si es una consecuencia a partir del conjunto vacío con respecto a D.

Suele exigirse que el conjunto de reglas de un sistema deductivo sea efectivamente decidible, y que no tomen en cuenta más que la *forma* —es decir, las características sintácticas— de las fórmulas del lenguaje formal en cuestión (es por ello que los lenguajes formales se llaman *formales*). Cuando éste sea el caso, existirá un algoritmo que nos permita determinar de cada secuencia finita de fórmulas si se trata de una prueba. En ese sentido, el conjunto de pruebas está al alcance de seres finitos como nosotros, al menos en principio.

Los lenguajes formales también suelen estar asociados con una *teoría de modelos* (ver sección 7.2). A partir de una teoría de modelos puede darse una caracterización matemática de las nociones intuitivas de *consecuencia lógica (o semántica)* y *verdad lógica* para fórmulas de F. Cuando un lenguaje formal está a la vez asociado con un sistema deductivo y con una teoría de modelos, pueden hacerse comparaciones entre la noción de *teorema* y la noción de *verdad lógica*. Si todos los teoremas de F con respecto a D son verdades lógicas se dice que D es un sistema deductivo *sólido* [*sound*, en inglés]. Cuando todas las verdades lógicas de F son teoremas con respecto a D se dice que D es un sistema deductivo *completo*.

2. FORMALIZACIÓN

Cuando formalizamos un enunciado del lenguaje ordinario lo parafraseamos utilizando un lenguaje formal, más apropiado para el desarrollo de nuestros objetivos lingüísticos. En la sección 2.1 hablaremos más sobre objetivos lingüísticos. En la sección 2.2 discutiremos la relación entre un lenguaje y su formalización.

2.1. *Objetivos lingüísticos*

Nuestros objetivos lingüísticos son las metas hacia las que están dirigidos nuestros usos del lenguaje. Por ejemplo, mi proferencia de 'tengo hambre' puede tener como objetivo que Julián me procure alimentos, y mi proferencia de '¡fuego!' puede tener como objetivo que Susana abandone apresuradamente el edificio.

Hay circunstancias en las que nuestros objetivos lingüísticos exigen características específicas de nuestro uso del lenguaje. Por ejemplo, en contextos ordinarios decir 'Susana tiene neumonía' es una buena manera de alcanzar el objetivo de que se sepa que Susana no va a ir a la fiesta. Pero hay circunstancias en las que no es así. Si estoy ante un grupo de niños pequeños, mi proferencia podría resultar poco efectiva porque los niños pequeños suelen no entender la palabra 'neumonía'. Para alcanzar mi objetivo, mi uso del lenguaje debe evitar palabras que sólo un adulto entendería. Por tanto, una *paráfrasis* resulta conveniente. En lugar de decir 'Susana tiene neumonía' digo 'Susana está muy enferma'. Para que la paráfrasis resulte apropiada lo único que hace falta es que, a mi juicio, me permita conseguir algo suficientemente cercano a mi objetivo lingüístico original.

A continuación se presenta una lista de características específicas que nuestros objetivos lingüísticos podrían exigir del lenguaje.

1) *Mi uso del lenguaje ha de evitar ambigüedades semánticas*
Ejemplo: Vengo de la mueblería y quiero hacerle saber a Julián que compré un banco. En otras circunstancias diría 'acabo de comprar un banco'. Pero Julián cree que soy una persona de muchísimo dinero, y mi proferencia le haría creer que adquirí una institución bancaria. Una paráfrasis resulta conveniente. Digo 'acabo de comprar un banco para sentarme'.

2) *Mi uso del lenguaje ha de incorporar ambigüedades semánticas*
Ejemplo: Regresé de la fiesta al amanecer, mientras Susana dormía. Al despertar, Susana me pregunta por la hora de mi llegada. Quiero con-

testar a su pregunta sin decir una falsedad. En otras circunstancias diría 'llegué de madrugada', pero sé que en este caso eso enojaría a Susana, y no quiero que se enoje. Una paráfrasis resulta conveniente. Digo 'llegué temprano'.

3) *Mi uso del lenguaje ha de evitar ambigüedades sintácticas*
Ejemplo: Se rumorea que las cosas están tan mal entre Julián y su pareja que pronto dejarán de estar juntos. Yo sé que los rumores son falsos y quiero hacérselo saber a Susana. En otras circunstancias diría 'No es el caso que Julián y su pareja están en malos términos y van a separarse'. Pero Susana tiene la teoría de que Julián y su pareja van a separarse a pesar de estar en buenos términos, e interpretaría mi proferencia como la conjunción de 'No es el caso que Julián y su pareja están en malos términos' y 'Julián y su pareja van a separarse'. Una paráfrasis resulta conveniente. Digo 'Ni están en malos términos Julián y su pareja ni van a separarse'.

4) *Mi uso del lenguaje ha de incorporar ambigüedades sintácticas*
Ejemplo: La madre de Susana está embarazada y Susana fue quien se lo dijo. Julián me pregunta por lo sucedido y no quiero comprometerme con una mentira. En otras circunstancias diría 'La madre de Susana está embarazada y Susana acaba de comunicárselo'. Pero estoy enojado con Julián, y quiero hacerle creer que la embarazada es Susana. Una paráfrasis resulta conveniente. Digo 'Susana acaba de decirle a su madre que está embarazada'.

5) *Mi uso del lenguaje ha de facilitar la comprensión de lo que digo*
Ejemplo: Estoy en un restaurante con Julián y creo que debe pedir pescado. En otras circunstancias diría 'Aquí sirven un pescado buenísimo'. Pero Julián sólo habla inglés. Una paráfrasis resultará conveniente. Digo: '*I recommend the fish*'.

6) *Mi uso del lenguaje ha de dificultar la comprensión de lo que digo*
Ejemplo: Quiero hacerle saber a Julián que la fiesta sorpresa de Susana será el sábado. En otras circunstancias diría 'La fiesta de Susana es el sábado'. Pero Susana está en la misma habitación y podría escucharnos. Una paráfrasis resulta conveniente. Digo 'Con respecto a lo que discutíamos el otro día, nos vemos el sábado'.

7) *Mi uso del lenguaje ha de incluir redundancias*
Ejemplo: Estoy hablando por teléfono con Julián y quiero hacerle saber que no habrá fiesta. En otras circunstancias diría 'decidimos pos-

poner la fiesta para la semana siguiente'. Pero Julián no puede escucharme bien porque nuestra conexión telefónica es muy mala. Una paráfrasis resulta conveniente. Digo 'No habrá fiesta. Quédate en casa'.

8) *Mi uso del lenguaje ha de ser breve*
Ejemplo: Le estoy escribiendo una carta a Susana y quiero hacerle saber que es un día muy hermoso. En otras circunstancias escribiría: 'Los jacarandás están en flor y no hay una sola nube en el cielo'. Pero estoy escribiendo un telegrama y tengo poco dinero. Una paráfrasis resulta conveniente. Digo: 'Buen tiempo'.

9) *Mi uso del lenguaje ha de resultar jocoso, o poético, o alegre, o triste, o cálido, o serio, o sombrío, o intimidador, o sereno, o grosero, o reconfortante, o solemne, o paternal, o respetuoso, etcétera.*
Ejemplo: Quiero hacerle saber a la concurrencia que Susana acaba de llegar. En otras circunstancias diría 'Susana está aquí'. Pero quiero que mi comentario resulte jocoso. Una paráfrasis resulta conveniente. Digo: '¡Hace su entrada la reina de las flores!'.

10) *He de utilizar un vocabulario básico restringido*
Ejemplo: Algunos filósofos creen que los números no existen. Si yo soy uno de ellos, y quiero hacerle saber a Susana que ni sobran ni faltan sillas, podría no querer decir 'el número de las sillas coincide con el número de los invitados', pues este enunciado involucra terminología numérica. Una paráfrasis resulta más conveniente. Digo: 'Hay tantas sillas como invitados'.

11) *He de utilizar un lenguaje que nos permita delimitar las consecuencias lógicas de mis proferencias*
Ejemplo: Algunos filósofos creen que la existencia de propiedades es problemática. Cuando digo: 'la nieve tiene la propiedad de ser blanca', podría no estar claro si me estoy comprometiendo con la existencia de propiedades, o si simplemente quiero decir 'la nieve es blanca'. Una paráfrasis resulta más conveniente. Digo: 'Existe un objeto x tal que x es una propiedad y la nieve tiene x'.

Ésta no ha sido, por supuesto, una lista exhaustiva.

2.2. *La relación entre un enunciado y su formalización*

Cuando formalizamos un enunciado lo parafraseamos utilizando un lenguaje formal, más apropiado para el desarrollo de nuestros objeti-

vos lingüísticos. La formalización es una herramienta, un medio para alcanzar nuestras metas. Por tanto, para que mis formalizaciones sean *correctas* lo único que hace falta es que sean *útiles*: que, a mi juicio, me permitan conseguir algo suficientemente cercano a mis objetivos lingüísticos originales[4]. En particular, para que una formalización sea correcta no hace falta que la paráfrasis sea *sinónimo* con el enunciado original, o que comparta su forma lógica o sus condiciones de verdad[5]. (Para que una formalización sea correcta, ni siquiera hace falta que el lenguaje formal en cuestión esté interpretado. Por ejemplo, mi objetivo podría consistir en la comprensión de la estructura sintáctica de un argumento.)

Hay casos especiales en donde sí requerimos sinonimia, o igualdad de forma lógica o igualdad de condiciones de verdad. Esto sucede cuando el desarrollo de nuestras metas requiere sinonimia, o igualdad de forma lógica o igualdad de condiciones de verdad. Por ejemplo, si formalizamos un argumento filosófico y queremos que nuestra formalización capture *exactamente* lo que dice el argumento original, podríamos requerir sinonimia, o algo cercano a la sinonimia. Casi siempre es difícil encontrar lenguajes formales que permitan satisfacer requerimientos tan severos, y estamos obligados a conformarnos con menos.

Las propiedades de una formalización no necesariamente son propiedades del enunciado original. Por ejemplo, al escoger una formalización no ambigua para un enunciado ambiguo, no eliminamos la ambigüedad en el enunciado original. Pero mientras estemos dispuestos a trabajar con la formalización, trabajaremos con una fórmula que no es ambigua.

En general, los lenguajes formales presentan las siguientes características:

1. Evitan ambigüedades sintácticas.
2. Sus fórmulas tienen una estructura sintáctica que es fácil de entender.
3. Sus fórmulas son breves.
4. Están asociados con interpretaciones que evitan ambigüedades semánticas y delimitan claramente los recursos expresivos del lenguaje.

4. Para este tipo de idea, véase Quine, 1960.

5. Raúl Orayen ha subrayado una consecuencia importante del hecho de que la formalización de un enunciado no necesariamente preserva su significado o su forma lógica: en general, no basta con formalizar un enunciado para poder aplicarle los criterios de consecuencia lógica y compromiso ontológico asociados al lenguaje formal en cuestión. Una excelente discusión sobre estos temas —y una respuesta de Quine a las observaciones de Orayen— pueden encontrarse en Orayen, 1989.

5. Están asociados con teorías de modelos que delimitan las consecuencias lógicas de sus fórmulas.
6. Están asociados con criterios que nos permiten delimitar los compromisos ontológicos de sus fórmulas.
7. Están asociados a sistemas deductivos y teorías de modelos con propiedades interesantes (en algunos casos, solidez y completud).

No siempre es buena idea formalizar. Todo depende de si nuestros objetivos requieren que nuestro lenguaje presente características como las de la lista anterior. Si, por ejemplo, nuestros objetivos requieren un lenguaje ambiguo, ameno o jocoso, formalizar puede ser mala idea. En otros casos, la formalización es una herramienta extremadamente valiosa.

3. LENGUAJES DE PRIMER ORDEN

Los lenguajes formales más frecuentemente utilizados en filosofía son los *lenguajes de primer orden*. Estos lenguajes son útiles porque presentan una capacidad expresiva sorprendentemente rica y combinan todas las características en la lista de la sección 2.2.

En la sección 3.1 se explican las características básicas de un lenguaje de primer orden. En la sección 3.2 se considera un método de traducción al castellano para los lenguajes de primer orden. En la sección 3.3 se presenta un sistema deductivo para lenguajes de primer orden.

3.1. *Caracterización básica*

Los símbolos básicos de un lenguaje de primer orden son: (*a*) para cada entero positivo *n*, una serie (finita o infinita) de letras de predicado: 'P_1^n', 'P_2^n', etc.; (*b*) el símbolo de identidad '='; (*c*) para cada entero positivo *n*, una serie (finita o infinita) de letras de función: 'f_1^n', 'f_2^n', etc.; (*d*) una serie (finita o infinita) de letras de constante: 'c_1', 'c_2', etc.; (*e*) una serie infinita de variables: 'x_1', 'x_2', etc.; (*f*) conectivas proposicionales básicas: '$\wedge$' y '$\neg$'; (*g*) el símbolo existencial: '$\exists$'; (*h*) paréntesis: '(' y ')'; e (*i*) la coma: ','.

Antes de dar las reglas de formación de un lenguaje de primer orden, es necesario definir la noción de *término*:

- Cualquier letra de constante es un término.
- Cualquier variable es un término.

- Si ⌜f_i^n⌝ es una letra de función y ⌜t_{j1}⌝, ..., ⌜t_{jn}⌝ son n términos, entonces ⌜$f_i^n(t_{j1}, ..., t_{jn})$⌝ es un término.
- Ninguna otra secuencia de símbolos básicos es un término.

Así, se sigue de nuestras definiciones que 'c_2', 'x_3', '$f_7^1(c_4)$', '$f_{16}^2(x_2,c_{15})$' y '$f_4^1(f_{16}^2(x_2,c_{15}))$' son términos. Teniendo la noción de término a bordo, las reglas de formación pueden formularse como sigue:

- Si ⌜P_i^n⌝ es una letra de predicado y ⌜t_{j1}⌝, ..., ⌜t_{jn}⌝ son n términos, entonces ⌜$P_i^n\ (t_{j1}, ..., t_{jn})$⌝ es una fórmula.
- Si ⌜ϕ⌝ y ⌜ψ⌝ son fórmulas, entonces ⌜$(\phi \wedge \psi)$⌝ es una fórmula.
- Si ⌜ϕ⌝ es una fórmula, entonces ⌜$(\neg\phi)$⌝ es una fórmula.
- Si ⌜t_i⌝ y ⌜t_j⌝ son términos, entonces ⌜$(t_i = t_j)$⌝ es una fórmula.
- Si ⌜ϕ⌝ es una fórmula y ⌜x_i⌝ es una variable, entonces ⌜$\exists x_i\ \phi$⌝ es una fórmula.
- Ninguna otra secuencia de símbolos básicos es una fórmula.

Así, se sigue de nuestras definiciones que '$P_4^1(x_3)$', '$P_7^2(c_2,f_{16}^2(x_2,c_{15}))$', '$x_3 = f_7^1(c_4)$', '$\exists x_4\ (x_3 = x_4)$', '$(P_4^1(x_3) \wedge (x_3 = x_4))$' y '$\exists x_4\ (\neg P_4^1(x_3))$' son fórmulas.

Además de conectivas proposicionales básicas, los lenguajes de primer orden suelen incluir conectivas proposicionales no básicas, definidas a partir de las básicas. '(A ∨ B)' suele entenderse como una abreviación de '(¬(¬(A) ∧ ¬(B)))', '(A ⊃ B)' suele entenderse como una abreviación de '(¬(A) ∨ B)', y '(A ≡ B)' suele entenderse como una abreviación de '((A ⊃ B) ∧ (B ⊃ A))'. También suelen incluir el símbolo universal '∀', definido a partir del existencial: ⌜$\forall x_i\ \phi$⌝ se entiende como una abreviación de ⌜$(\neg\exists x_i\ (\neg\phi))$⌝. En formulaciones distintas de lenguajes de primer orden, otras conectivas son consideradas como básicas y '∧' o '¬' (o ambas) se toman como derivadas. Algunas veces se considera al símbolo universal como símbolo básico y al existencial como derivado.

Algunas definiciones serán útiles más adelante. En la fórmula ⌜$\exists x_i\ \phi$⌝ se dice que ⌜ϕ⌝ es el *rango* del cuantificador ⌜$\exists x_i$⌝. Se dice que la ocurrencia de una variable ⌜x_i⌝ en una fórmula ϕ es *ligada* si y sólo si es parte de un cuantificador ⌜$\exists x_i$⌝ o se encuentra dentro del rango de un cuantificador ⌜$\exists x_i$⌝. En caso contrario se dice que la ocurrencia de la variable es *libre* en la fórmula. Si ϕ es una fórmula y t es un término, se dice que t está libre para x en ϕ si y sólo si ninguna ocurrencia libre de x en ϕ está en el rango de un cuantificador ⌜$\exists x_i$⌝ donde ⌜x_i⌝ es una variable en t. Finalmente, decimos que una fórmula es un *enunciado* si no contiene variables libres.

3.2. *Un método de traducción al castellano para los lenguajes de primer orden*

Las fórmulas de un lenguaje de primer orden pueden ser traducidas al castellano. El primer paso consiste en decidir de qué objetos queremos que hable nuestro lenguaje. Decimos que el *dominio* del lenguaje consiste de esos objetos. El segundo paso consiste en asignar traducciones a las letras de predicado, función y constante. Por ejemplo, podríamos traducir 'P_1^1' como '... es griego', 'P_1^2' como '... y ... son tales que el primero es más valiente que el segundo', 'f_1^1' como 'el padre de ...', 'c_1' como 'Sócrates' y 'c_2' como 'Julio César'. Finalmente, extendemos nuestra traducción a las fórmulas del lenguaje como sigue:

- Si ϕ es una fórmula de la forma $\ulcorner(\neg\psi)\urcorner$, entonces la traducción de ϕ es $\ulcorner$no es el caso que $\psi^{tr}\urcorner$ (donde $\ulcorner\psi^{tr}\urcorner$ es la traducción de $\ulcorner\psi\urcorner$).
- Si ϕ es una fórmula de la forma $\ulcorner(\psi \wedge \xi)\urcorner$, entonces la traducción de ϕ es $\ulcorner$es el caso tanto que ψ^{tr} como que $\xi^{tr}\urcorner$ (donde $\ulcorner\psi^{tr}\urcorner$ es la traducción de $\ulcorner\psi\urcorner$ y $\ulcorner\xi^{tr}\urcorner$ es la traducción de $\ulcorner\xi\urcorner$).
- Si ϕ es una fórmula de la forma $\ulcorner\exists x_i\ \psi\urcorner$, entonces la traducción de ϕ es $\ulcorner$existe un objeto$_i$ en el dominio tal que $\psi^{tr}\urcorner$ (donde $\ulcorner\psi^{tr}\urcorner$ es la traducción de $\ulcorner\psi\urcorner$).
- La traducción de una variable $\ulcorner x_i\urcorner$ es $\ulcorner$ese$_i$ objeto$\urcorner$.
- Si ϕ es una fórmula de la forma $\ulcorner(t_i = t_j)\urcorner$, entonces la traducción de ϕ es $\ulcorner t_i^{tr}$ es idéntico a $t_j^{tr}\urcorner$ (donde $\ulcorner t_i^{tr}\urcorner$ es la traducción de $\ulcorner t_i\urcorner$ y $\ulcorner t_j^{tr}\urcorner$ es la traducción de $\ulcorner t_j\urcorner$).
- Si ϕ es una fórmula de la forma $\ulcorner P_i^n\ (t_{j1}, ..., t_{jn})\urcorner$, entonces la traducción de ϕ es $\ulcorner t_{j1}^{tr}$ y ... y $t_{jn}^{tr}\ (P_i^n)^{tr}\urcorner$ (donde $\ulcorner(P_i^n)^{tr}\urcorner$ es la traducción de $\ulcorner P_i^n\urcorner$ y para cada $k \leq n$, $\ulcorner t_{jk}^{tr}\urcorner$ es la traducción de $\ulcorner t_{jk}\urcorner$).
- Si ϕ es un término de la forma $\ulcorner f_i^n\ (t_{j1}, ..., t_{jn})\urcorner$, entonces la traducción de ϕ es $\ulcorner(f_i^n)^{tr}\ t_{j1}^{tr}$ y . . . y $t_{jn}^{tr}\urcorner$ (donde $\ulcorner(f_i^n)^{tr}\urcorner$ es la traducción de $\ulcorner f_i^n\urcorner$ y para cada $k \leq n$, $\ulcorner t_{jk}^{tr}\urcorner$ es la traducción de $\ulcorner t_{jk}\urcorner$).

He aquí algunos ejemplos de nuestra traducción:

'$P_1^1(c_1)$'	⇨	'Sócrates es griego'.
'$c_1 = c_2$'	⇨	'Sócrates es idéntico a Julio César'.
'$f_1^1\ (c_1)$'	⇨	'el padre de Sócrates'.
'$P_1^2(f_1^1\ (c_1),c_2)$'	⇨	'el padre de Sócrates y Julio César son tales que el primero es más valiente que el segundo'.

'$(\neg P_1^1(c_1))$'	⇨	'no es el caso que Sócrates es griego'.
'$(P_1^1(c_1) \wedge P_1^1(c_2))$'	⇨	'es el caso tanto que Sócrates es griego como que Julio César es griego'.
'$P_1^1\ (x_1)$'	⇨	'ese_1 objeto es griego'.
'$\exists x_1\ P_1^1\ (x_1)$'	⇨	'existe un $objeto_1$ en el dominio tal que ese $objeto_1$ es griego'.
'$\exists x_1 \exists x_2\ P_1^2\ (x_1,x_2))$'	⇨	'existe un $objeto_1$ en el dominio tal que existe un $objeto_2$ en el dominio tal que ese_1 objeto y ese_2 objeto son tales que el primero es más valiente que el segundo'.

Nuestro método de traducción produce enunciados tediosos y poco elegantes, pero significativos. Los subíndices en la traducción se utilizan para evitar ambigüedades. Por ejemplo, en la ausencia de subíndices, la traducción de la última fórmula en la lista precedente resultaría ambigua. En lugar de subíndices podríamos utilizar frases como 'éste', 'aquél', 'el primero', 'el segundo', 'el tercero', etc. Pero esto haría las traducciones todavía más tediosos. La presencia de subíndices implica que las traducciones no son, estrictamente hablando, traducciones al castellano. Pero no por ello dejan de ser significativas. Pueden ser comprendidas por cualquier persona que conozca el castellano.

Nuestra traducción le asigna a cualquier *enunciado* del lenguaje formal (cualquier fórmula sin variables libres) un enunciado verdadero o falso del castellano con subíndices. Pero hay fórmulas a las que nuestra traducción asigna enunciados del castellano que no son ni verdaderos ni falsos. Por ejemplo, nuestra traducción le asigna 'ese_1 objeto es griego' a '$P_1^1\ (x_1)$'. Dado que no sabemos quién es 'ese_1 objeto', no podemos saber si 'ese_1 objeto es griego' es verdadero o falso. Es ilustrativo comparar este resultado con el resultado análogo de la sección 5.1.

3.3. *Sistemas deductivos para lenguajes de primer orden*

Un lenguaje de primer orden puede estar asociado con diferentes sistemas deductivos, pero muchos de ellos son equivalentes en el sentido de que precisamente las mismas fórmulas resultan ser teoremas. Un sistema deductivo de primer orden es un sistema *clásico* si arroja los mismos teoremas que el que se presenta a continuación (Mendelson, 1997):

- Si está permitido derivar ⌜(A ⊃ B)⌝ y ⌜A⌝ a partir del conjunto P, entonces también está permitido derivar ⌜B⌝ a partir del conjunto P.
- Si está permitido derivar ⌜A⌝ a partir del conjunto P, entonces también está permitido derivar ⌜$\forall x_i$ A⌝ a partir del conjunto P.
- Está permitido derivar cualquier fórmula ⌜(A ⊃ (B ⊃ A))⌝ a partir de cualquier conjunto P[6].
- Está permitido derivar cualquier fórmula ⌜((A ⊃ (B ⊃ C)) ⊃ ((A ⊃ B) ⊃ (A ⊃ C)))⌝ a partir de cualquier conjunto P.
- Está permitido derivar cualquier fórmula ⌜((¬B ⊃ ¬A) ⊃ ((¬B ⊃ A) ⊃ B))⌝ a partir de cualquier conjunto P.
- Está permitido derivar cualquier fórmula ⌜$\forall x_i$ A ⊃ A[t/x_i]⌝ a partir de cualquier conjunto P, (siempre y cuando ⌜A[t/x_i]⌝ sea el resultado de reemplazar cada ocurrencia de ⌜x_i⌝ por el término ⌜t⌝ en ⌜A⌝ y ⌜t⌝ esté libre para ⌜x_i⌝ en ⌜A⌝).
- Está permitido derivar cualquier fórmula ⌜$\forall x_i$ ((A ⊃ B) ⊃ (A ⊃ $\forall x_i$ B))⌝ a partir de cualquier conjunto P, (siempre y cuando ⌜A⌝ no contenga ocurrencias libres de ⌜x_i⌝).
- Si ϕ es un miembro de P, está permitido derivar ϕ a partir de P.

En la sección 7.1 se desarrolla una teoría de modelos para lenguajes de primer orden. Con respecto a ella, los sistemas deductivos clásicos son sólidos y completos.

4. EJEMPLOS DE FORMALIZACIONES

En este apartado formalizaremos algunos enunciados del castellano utilizando un lenguaje de primer orden. Comencemos con el siguiente argumento:

Las ballenas son mamíferos
Bernardo es una ballena

Bernardo es un mamífero.

En muchos casos, la siguiente es una buena formalización de este argumento a un lenguaje de primer orden:

6. O, equivalentemente, si está permitido derivar todas las fórmulas en el conjunto vacío a partir de un conjunto cualquiera P, entonces también está permitido derivar ⌜(A → (B → A))⌝ a partir de P.

$\forall x$ (Ballena(x) $\supset$ Mamífero(x))
Ballena(Bernardo)

Mamífero(Bernardo).

Al formalizar, hemos utilizado 'Ballena' y 'Mamífero' como letras de predicado, en lugar de utilizar, por ejemplo, 'P_1^1' y 'P_2^1'. También utilizamos 'Bernardo' como letra de constante en lugar de utilizar, por ejemplo, 'c_1', y 'x' como variable en lugar de utilizar, por ejemplo, 'x_1'. Este tipo de laxitud —común entre filósofos y científicos— facilita enormemente la comprensión de nuestros enunciados y suele no interferir con nuestros objetivos. La pondremos en práctica durante el resto del artículo.

Nuestra formalización tiene algunas propiedades de las que el argumento original carece. Por ejemplo, es fácil verificar que existe una *prueba* en un sistema deductivo clásico de la conclusión del argumento formalizado a partir de sus premisas. En cambio, la noción de prueba no está bien definida para el argumento original, porque no está bien definido para el lenguaje ordinario.

Vale la pena notar también que no está claro que nuestra formalización de 'las ballenas son mamíferos' preserve su significado. De acuerdo con la traducción al castellano de la sección 3.2, '$\forall x$ (Ballena(x) $\supset$ Mamífero(x))' significa algo equivalente a 'si algo es una ballena entonces es un mamífero'. Pero no es obvio que 'las ballenas son mamíferos' signifique lo mismo que 'si algo es una ballena entonces es un mamífero'. Lo primero podría ser interpretado como una afirmación acerca del tipo *ballena*, y lo segundo no.

También vale la pena notar que un argumento como el anterior no puede ser formalizado en un cálculo proposicional si requerimos que el resultado sea lógicamente válido.

Consideremos otro ejemplo. El enunciado 'Todas los chicos están enamorados de una chica' se presta a las dos formalizaciones distintas:

(A) $\forall x$ (Chico(x) $\supset \exists y$ (Chica(y) $\wedge$ Enamorado(x, y)))

(B) $\exists y$ (Chica(y) $\wedge$ $\forall x$ (Chico(x) $\supset$ Enamorado(x, y)))

La primera fórmula puede leerse como 'para todo chico existe alguna chica de la que el chico está enamorado'. La segunda puede leerse como 'hay una chica tal que todos los chicos están enamorados de ella'. El enunciado castellano original es ambiguo entre las dos inter-

pretaciones. Al formalizarlo como (A) o formalizarlo como (B) no eliminamos la ambigüedad en el enunciado original. Pero en tanto que estemos dispuestos a trabajar con las formalizaciones, trabajaremos con enunciados que no son ambiguos.

5. LA INTERPRETACIÓN DE UN LENGUAJE FORMAL

Se dice que un lenguaje formal está interpretado cuando sus fórmulas han recibido una asignación de significado. Frecuentemente, la manera más fácil de asignar significados es proporcionando una *traducción* a un lenguaje previamente significativo —un lenguaje natural, por ejemplo—. Consideramos un ejemplo en la sección 3.2.

En este apartado discutiremos otros métodos. En la sección 5.1 se habla de las semánticas formales, y en la subsección 5.1.1 del sentido en el que puede decirse que una semántica formal resulta en una asignación de significados. En la sección 5.2 se considera un método axiomático. En la sección 5.3 se hace una breve comparación entre los distintos métodos.

5.1. *Semánticas formales*

Una *semántica formal* para un lenguaje formal F es una asignación de valores de verdad a un subconjunto de fórmulas de F.

Se dice que una semántica formal para F es *composicional* siempre y cuando (*a*) especifique un conjunto de secuencias de símbolos básicos que han de contar como *unidades significativas básicas*, (*b*) le asigne valores semánticos a las unidades significativas básicas de F, (*c*) le asigne un valor semántico a cada fórmula de F como función de los valores semánticos de las unidades significativas básicas que aparecen en la fórmula, junto con la estructura sintáctica de la fórmula, y (*d*) le asigne valores de verdad a un subconjunto de fórmulas de F como función de sus valores semánticos[7].

Como ejemplo, consideremos una semántica formal para un lenguaje de primer orden L. Las unidades significativas básicas de nuestra semántica formal son las variables, las letras de predicado, función y constante, las conectivas proposicionales, el símbolo existencial '∃' y el símbolo de identidad '='. Comenzamos escogiendo una clase D, que jugará el papel de nuestro dominio, y decimos que el valor se-

7. Una semántica formal composicional para lenguajes de primer orden fue desarrollada por primera vez en Tarski, 1933.

mántico de cualquier variable es D. Para cada letra de predicado $\ulcorner P_i^n \urcorner$ escogemos como valor semántico una clase de *n*-tuplos ordenados de elementos de D. Para cada letra de función $\ulcorner f_i^n \urcorner$ escogemos como valor semántico una función de *n*-tuplos ordenados de elementos de D a elementos de D. Para cada letra de constante $\ulcorner c_i \urcorner$ escogemos como valor semántico un elemento de D. Para el símbolo de identidad '=' escogeremos como valor semántico la clase de todos los pares ordenados $<x, x>$.

Antes de escoger valores semánticos para las unidades significativas básicas restantes, es necesario definir la noción de *valuación*. Una valuación v es una función de variables a objetos. Como valor semántico de la conectiva proposicional '$\neg$' escogemos la función que nos lleva de cada clase de valuaciones V a la clase de valuaciones que no están en V. Como valor semántico de la conectiva proposicional '$\wedge$' escogemos la función que nos lleva de las clases de valuaciones V y W a la intersección de V y W. Finalmente, como valor semántico para '$\exists$' escogemos la (súper) clase de todas las clases no vacías[8].

La asignación de valores semánticos a fórmulas arbitrarias de L requiere de dos nociones preliminares: *denotación* y *satisfacción*. La *denotación* es una función, $d(v, t)$, que va de cada valuación v y término t a un elemento del dominio. Puede definirse como sigue:

- Si t es una variable, $d(v, t) = v(\mathrm{x})$.
- Si t es una letra de constante, $d(v, t)$ es el valor semántico de t.
- Si t es $\ulcorner \mathrm{f}_i^n(s_{j1}, ..., s_{jn}) \urcorner$, donde $\ulcorner s_{j1} \urcorner$, ..., $\ulcorner s_{jn} \urcorner$ son términos, $d(v, t)$ es el resultado de aplicar el valor semántico de 'f_i^n' a $<v(s_{j1}), ..., v(s_{jn})>$.

La *satisfacción* es una relación entre una valuación v y una fórmula ϕ, y puede definirse como sigue:

- Si ϕ tiene la forma $\ulcorner (s_i = s_j) \urcorner$, donde $\ulcorner s_i \urcorner$ y $\ulcorner s_j \urcorner$ son términos, v satisface a ϕ si y sólo si $< d(v, \ulcorner s_i \urcorner), d(v, \ulcorner s_j \urcorner)>$ es un miembro del valor semántico de '='.
- Si ϕ tiene la forma $\ulcorner P_i^n(s_{j1}, ..., s_{jn}) \urcorner$, donde $\ulcorner s_{j1} \urcorner$, ..., $\ulcorner s_{jn} \urcorner$ son términos, v satisface a ϕ si y sólo si $< d(v, \ulcorner s_{j1} \urcorner), ..., d(v, \ulcorner s_{jn} \urcorner)>$ es un miembro del valor semántico de $\ulcorner P_i^n \urcorner$.

8. Para evitar las paradojas de la teoría de conjuntos es necesario asumir que las clases no son objetos y que las súperclases no son ni objetos ni clases. Una manera de lograr esto es tratando la cuantificación sobre clases como cuantificación de segundo orden y la cuantificación sobre súper-clases como cuantificación de tercer orden. Véase Boolos, 1998, especialmente la primera sección, y Rayo y Yablo, 2000.

- Si ϕ tiene la forma $\ulcorner(\psi \wedge \xi)\urcorner$, v satisface a ϕ si y sólo si v está en la clase de valuaciones que resulta de aplicar el valor semántico de '$\wedge$' a V y W, donde V es la clase de valuaciones que satisfacen a $\ulcorner\psi\urcorner$ y W es la clase de valuaciones que satisfacen a $\ulcorner\xi\urcorner$.
- Si ϕ tiene la forma $\ulcorner(\neg\psi)\urcorner$, v satisface a ϕ si y sólo si v está en la clase de valuaciones que resulta de aplicar el valor semántico de '$\neg$' a la clase de valuaciones que satisfacen a $\ulcorner\psi\urcorner$.
- Si ϕ tiene la forma $\ulcorner\exists \mathrm{x}_i\ \psi\urcorner$, v satisface a ϕ si y sólo si X_i es la intersección del valor semántico de $\ulcorner \mathrm{x}_i \urcorner$ con algún miembro del valor semántico de '$\exists$', donde X_i es la clase de objetos z tal que el resultado de modificar a v para que la asignación de $\ulcorner \mathrm{x}_i \urcorner$ sea z satisface a $\ulcorner\psi\urcorner$.

Finalmente, decimos que el valor semántico de una fórmula de L es la clase de secuencias que la satisfacen. Decimos que una fórmula de L es *verdadera* si y sólo si su valor semántico contiene todas las secuencias, y decimos que es *falsa* si y sólo si su valor semántico es la clase vacía. Es fácil verificar que nuestra semántica formal es composicional, en el sentido proporcionado arriba.

Para fijar las ideas, consideremos un ejemplo más concreto. Digamos que el valor semántico de $\ulcorner \mathrm{x}_i \urcorner$—es decir, el dominio— es la clase de todas las personas, y que el valor semántico de 'P_1^1' es la clase de personas griegas.

Consideremos la fórmula '$P_1^1(x_1)$'. De acuerdo con nuestra definición, una valuación v satisface a '$P_1^1(x_1)$' si y sólo si $d(v, \text{'}x_1\text{'})$ es un miembro del valor semántico de 'P_1^1', es decir, la clase de las personas griegas. Dado que 'x_1' es una variable, $d(v, \text{'}x_1\text{'}) = v(\text{'}x_1\text{'})$. Por tanto, v satisface a '$P_1^1(x_1)$' si y sólo si $v(\text{'}x_1\text{'})$ es miembro del conjunto de las personas griegos. De esto se sigue que diferentes selecciones de v arrojan diferentes respuestas a la pregunta de si v satisface a '$P_1^1(x_1)$'. Existen valuaciones que satisfacen a '$P_1^1(x_1)$' y valuaciones que no la satisfacen. Por tanto, de acuerdo con nuestra semántica formal, la fórmula '$P_1^1(x_1)$' no es ni verdadera ni falsa.

Consideremos ahora un caso que involucre cuantificación, por ejemplo '$\exists x_1\ P_1^1(x_1)$'. Se sigue de nuestras definiciones que una secuencia v satisface a '$\exists x_1\ P_1^1(x_1)$' si y sólo si X_i es una subclase no vacía del dominio, donde X_i es la clase de objetos z tal que el resultado de modificar a v para que la asignación de $\ulcorner x_i \urcorner$ sea z satisface a $\ulcorner\psi\urcorner$. Pero es una consecuencia del ejemplo anterior, y del hecho de que nuestro dominio incluye griegos y no griegos, que X_i es no vacía. Por tanto v satisface a '$\exists x_1\ P_1^1(x_1)$'. Dado que nuestra selección de v fue arbitraria, se sigue que toda asignación satisface a '$\exists x_1$

$P_1{}^1(x_1)$'. Así, de acuerdo con nuestra semántica formal,'$\exists x_1\ P_1{}^1(x_1)$' es verdadera.

Finalmente, consideremos un caso que involucre conectivas proposicionales, por ejemplo, '$(\neg\exists x_1\ P_1{}^1(x_1))$'. De acuerdo con nuestra definición, una secuencia v satisface a '$(\neg\exists x_1\ \mathrm{P}_1{}^1(x_1))$' si y sólo si v está en la clase de valuaciones que no satisfacen a $\ulcorner\psi\urcorner$. Pero en el ejemplo anterior vimos que toda secuencia satisface a '$\exists x_1\ P_1{}^1(x_1)$'. Por tanto, v no satisface a '$(\neg\exists x_1\ P_1{}^1(x_1))$'. Dado que nuestra selección de v fue arbitraria, se sigue que ninguna asignación satisface a '$(\neg\exists x_1\ P_1{}^1(x_1))$'. Así, de acuerdo con nuestra semántica formal, '$(\neg\exists x_1\ P_1{}^1(x_1))$' es falsa.

Es posible probar que, de acuerdo con nuestra semántica formal, cualquier *enunciado* (cualquier fórmula sin variables libres) es verdadero o falso, pero hemos visto que hay fórmulas con variables libres que no son ni verdaderas ni falsas, como '$P_1{}^1(x_1)$'. Es ilustrativo comparar este resultado con el resultado análogo de la sección 3.2.

5.1.1. Asignación de significados por medio de una semántica formal

Cuando un lenguaje formal es interpretado por medio de una traducción a un lenguaje previamente significativo no hay ninguna duda que el resultado es una asignación de significados. Podemos decir simplemente que el significado de una fórmula es el significado de su traducción. ¿Pero en qué sentido puede decirse que una semántica formal para un lenguaje formal F proporciona una asignación de significados a las fórmulas de F?

La extensión de un predicado 'P' (es decir, la clase de objetos de los que 'P' es verdadero) nos proporciona cierta información acerca de su significado. En particular, sabemos que predicados con extensiones diferentes tienen significados diferentes. Por ejemplo, podemos establecer que los predicados castellanos 'es una fruta' y 'es un lagarto' tienen significados diferentes, señalando que sus extensiones no son coextensivas. Pero esto no significa que la extensión de un predicado nos proporcione información *completa* sobre su significado. Para ver esto basta con notar que los predicados castellanos 'es un animal cordado' y 'es un animal renado' tienen significados diferentes pero comparten la misma extensión.

Retornemos ahora a la semántica formal para lenguajes de primer orden de la sección 5.2. Podemos pensar en cada letra de predicado como un predicado, y en el valor semántico que le proporciona nuestra semántica formal como su extensión. De acuerdo con el párrafo anterior esto no nos proporciona información completa sobre el sig-

nificado del predicado, pero sí nos proporciona cierta información. En este sentido, puede decirse que la asignación de un valor semántico para $\ulcorner P_i^n \urcorner$ que efectúa nuestra semántica formal constituye una asignación *parcial* de significado. Algo similar puede decirse en el caso de las letras de función, las letras de constante, las conectivas proposicionales, el símbolo existencial y el símbolo de identidad. Por ejemplo, podemos pensar en cada letra de constante como un término singular, y en el valor semántico que le proporciona nuestra semántica formal como su referente.

Si concedemos que la asignación de valor semántico a las unidades significativas básicas de nuestro lenguaje de primer orden constituye una asignación parcial de significado, hay buenas razones para pensar que la asignación de valores semánticos a fórmulas arbitrarias también constituye una asignación parcial de significados. Así como los significados de 'Sócrates' y 'es griego' determinan el significado de 'Sócrates es griego' en castellano, una asignación (parcial) de significados a 'c_1' y 'P_1^1' puede determinar (de manera parcial) el significado de '$P_1^1(c_1)$'.

El hecho de que una asignación de significados sea *parcial* no quiere decir que sea *insuficiente*. Cuán cabalmente determinados deben estar los significados de nuestras fórmulas depende de cuáles sean nuestros objetivos, y no todos nuestros objetivos requieren de una determinación significativa total.

5.2. *El método axiomático*

Una manera diferente de asignarle significados a las fórmulas de un lenguaje formal es a través de axiomas o reglas axiomáticas.

Consideremos primero un ejemplo que no involucre lenguajes formales. Estipulemos que el término singular 'Clat' ha de ser usado de tal manera que el siguiente axioma resulte verdadero:

Clat es la persona que ahora se encuentra en la habitación 104.

Por sí sola, nuestra estipulación no garantiza que el axioma es verdadero. Supongamos, por ejemplo, que la habitación 104 resulta estar vacía, o que hay más de una persona dentro de ella. En ese caso, la estipulación no habrá sido exitosa en hacer de 'Clat' una expresión significativa. Pero supongamos que Julián es la única persona que se encuentra en la habitación 104. En ese caso es razonable pensar que, como resultado de nuestra estipulación, el término 'Clat' refiere a Julián.

Algunos filósofos creen que un método similar puede utilizarse en el caso de los lenguajes formales. Podríamos, por ejemplo, estipular que la conectiva proposicional '∧' ha de ser utilizada de tal modo que cada instancia de las siguientes reglas axiomáticas resulte lógicamente válida[9]:

$$\frac{\phi \wedge \psi}{\phi} \qquad \frac{\phi \wedge \psi}{\psi} \qquad \frac{\{\phi, \psi\}}{\phi \wedge \psi}$$

Dado que la conjunción es la única operación que puede hacer estas tres reglas lógicamente verdaderas, es razonable pensar que, como resultado de nuestra estipulación, la conectiva '∧' adquiere la conjunción como significado.

Gerhard Gentzen mostró que reglas de este tipo pueden utilizarse para delimitar el significado de las conectivas proposicionales. Resultados más recientes muestran que los resultados de Gentzen pueden extenderse más aún[10].

Vale la pena señalar que no cualquier estipulación como la que utilizamos para asignarle un significado a '∧' es exitosa. Utilizando un famoso ejemplo de Arthur Prior (1961), estipulemos que la conectiva proposicional 'tonk' ha de ser utilizada de tal modo que cada instancia de las siguientes reglas axiomáticas resulte lógicamente válida:

$$\frac{\phi \text{ tonk } \psi}{\phi} \qquad \frac{\phi \text{ tonk } \psi}{\psi} \qquad \frac{\phi}{\phi \text{ tonk } \psi} \qquad \frac{\psi}{\phi \text{ tonk } \psi}$$

Ninguna operación puede satisfacer estas reglas. Si existiera alguna estaríamos obligados a concluir que cualquier fórmula es una consecuencia lógica de cualquier otra. (*Prueba*: sean P y Q dos fórmulas cualesquiera. De acuerdo con la tercera regla, P tonk Q es una consecuencia lógica de P; de acuerdo con la segunda regla, Q es una consecuencia lógica de P tonk Q; por la transitividad de la relación de consecuencia lógica, Q es una consecuencia lógica de P. QED.) El proyecto de identificar conjuntos de reglas axiomáticas que permiten estipulaciones exitosas no es trivial.

El método axiomático también puede utilizarse para asignar significados a las letras proposicionales, de predicado, de función y de constante de un lenguaje de primer orden. Por ejemplo, podríamos

9. Aquí 'lógicamente válida' ha de ser entendido en su sentido preteorético.

10. Los resultados seminales pueden encontrarse en Gentzen, 1969. Véase también Koslow, 1992, y McGee, 1999.

estipular que las letras de predicado 'punto', 'línea recta', etc., sean utilizadas de tal manera que todos los axiomas de la geometría de Hilbert resulten verdaderos. Algunos filósofos piensan que esta estipulación resultaría en una asignación de significados.

5.3. *Comparación*

Hemos considerado tres métodos distintos para asignarle significados a las fórmulas de un lenguaje formal: el método de la traducción (sección 3.2), el método de la semántica formal (sección 5.1) y el método axiomático (sección 5.2).

Cuál de los distintos métodos ha de ser utilizado depende de nuestros objetivos. El método de la traducción, a diferencia de los otros dos, tiene la ventaja de hacer las fórmulas del lenguaje formal en cuestión inmediatamente comprensibles para un usuario del lenguaje al que se traduce. El método de la semántica formal, a diferencia de los otros dos, tiene la ventaja de proporcionar herramental matemático que resulta extraordinariamente útil para el estudio de algunas de las propiedades de los lenguajes formales. El método axiomático, a diferencia de los otros dos, tiene la ventaja de no requerir la comprensión previa de los significados que se asignan en la interpretación.

Es importante notar, sin embargo, que los distintos métodos no necesariamente están en competencia. Podemos utilizar distintos métodos para interpretar diferentes aspectos de un lenguaje formal. O podemos utilizar versiones de más de un método a la vez, cuidando que la interpretación que provenga de uno no esté en conflicto con las interpretaciones procedentes de los demás.

6. EXTENSIONES DE PRIMER ORDEN

En este apartado veremos que los lenguajes de primer orden tienen ciertas limitaciones expresivas, y consideraremos dos maneras de enfrentar la situación. En la sección 6.1 consideramos lenguajes modales. En la sección 6.2 consideramos lenguajes de segundo orden.

6.1. *Lenguajes modales*

En un lenguaje de primer orden no hay ninguna manera directa de expresar operadores como 'necesariamente' en 'necesariamente, el agua es H_2O' o 'Julián sabe que' en 'Julián sabe que el agua es H_2O'.

Esto no significa que sea imposible formalizar enunciados conteniendo estos operadores en un lenguaje de primer orden. Por ejemplo 'Necesariamente, el agua es H_2O' puede ser formalizado como 'Necesario(*c*)', donde el predicado 'Necesario (...)' es interpretado como '... es una proposición necesariamente verdadera', y la letra de constante '*c*' refiere a la proposición de que el agua es H_2O. Pero este tipo de formalización tiene desventajas. Por ejemplo, 'el agua es H_2O' es una consecuencia inmediata de 'Necesariamente, el agua es H_2O'. Pero, en un lenguaje de primer orden, no hay ninguna conexión inmediata entre 'Necesario(*c*)' y 'el agua = H_2O' (donde 'el agua' refiere al agua y 'H_2O' refiere al H_2O)[11].

Para solucionar esta limitación expresiva, pueden utilizarse versiones enriquecidas de lenguajes de primer orden llamados lenguajes *modales*. El enriquecimiento consiste en añadir un símbolo operacional '□' a los símbolos básicos del lenguaje, e incorporar la siguiente regla de formación:

- Si ⌜ϕ⌝ es una fórmula, entonces ⌜□ ϕ⌝ es una fórmula.

Una vez enriquecido el lenguaje, podemos interpretar '□' como 'necesariamente, ...', y expresar 'necesariamente, el agua es H_2O' utilizando la fórmula '□ (el agua = H_2O)'. Los lenguajes modales con esta interpretación son enormemente útiles en filosofía[12].

Interpretaciones distintas de '□' dan como resultado lenguajes modales con capacidades expresivas distintas. Por ejemplo, cuando '□' se interpreta como 'Julián sabe que ... ', obtenemos un lenguaje modal epistémico que nos permite formalizar 'Juan sabe que el agua es H_2O' como '□(el agua = H_2O)' (Hintika e Hintika, 1989); cuando '□' se interpreta como 'puede probarse que ...' obtenemos un lenguaje modal de las pruebas que nos permite formalizar 'puede probarse que 2 + 2 = 4' como '□ (2 + 2 = 4)' (Boolos, 1993).

11. Vale la pena señalar que el problema no puede solucionarse añadiendo un predicado de verdad 'verdadero(...)' al lenguaje, e insistiendo que 'el agua = H_2O' se sigue de 'Verdadero(*c*)', que a su vez se sigue de 'Necesario(*c*)'. Tarski probó que cualquier teoría (suficientemente fuerte) en la que pueda probarse cada enunciado que resulte de reemplazar el nombre de una proposición por '*d*' y un enunciado expresando esa proposición por '*x*' en 'Verdadero(*d*) ≡ *x*' es inconsistente.

12. Para una exposición de las características formales de este tipo de lenguajes, véase Chellas, 1980. Para una exposición acerca de la importancia filosófica de este tipo de lenguajes ver Lewis, 1986.

6.2. *Lenguajes de segundo orden*

La teoría de conjuntos estándar, ZFC[13], suele formularse en un lenguaje de primer orden, y cada instancia del llamado esquema de comprensión se considera como un axioma:

$$\exists y\ \forall x\ (x \in y \equiv (x \in z \wedge \mathcal{P}(x))$$

(donde $\ulcorner \mathcal{P}(x) \urcorner$ es reemplazado por una fórmula cualquiera con la variable $\ulcorner x \urcorner$ libre). Como resultado, obtenemos un conjunto infinito de axiomas. Juntos, los axiomas expresan algo parecido a la idea intuitiva de que el esquema de comprensión es válido para cualquier propiedad $\ulcorner \mathcal{P}(x) \urcorner$ *que puede ser expresada por una fórmula del lenguaje*. Pero los teóricos de conjuntos quisieran tener más. Quisieran poder expresar algo parecido a la idea intuitiva de que el esquema de comprensión es válido para *cualquier* propiedad $\ulcorner \mathcal{P}(x) \urcorner$, sea o no expresada por una fórmula del lenguaje. Desdichadamente, los recursos expresivos de un lenguaje de primer orden no nos permiten hacerlo.

Como remedio, puede utilizarse un lenguaje de *segundo orden*. Un lenguaje de segundo orden es el resultado de extender los símbolos básicos de un lenguaje de primer orden con variables de segundo orden: 'X_1', 'X_2', ..., y añadir las siguientes reglas de formación:

- Si $\ulcorner \phi \urcorner$ es una fórmula y $\ulcorner X_i \urcorner$ es una variable de segundo orden, entonces $\ulcorner \exists X_i\ \phi \urcorner$ es una fórmula.
- Si $\ulcorner t \urcorner$ es un término y $\ulcorner X_i \urcorner$ es una variable de segundo orden, entonces $\ulcorner X_i(t) \urcorner$ es una fórmula.

Así como en un lenguaje de primer orden podemos reemplazar la letra de constante 'Susana' en 'Corre(Susana)' por una variable, obteniendo 'Corre(x_1)', en un lenguaje de segundo orden podemos reemplazar la letra de *predicado* 'Corre' en 'Corre(Susana)', obteniendo 'X_1(Susana)'. Y así como en un lenguaje de primer orden podemos anclar la variable 'x_1' en 'Corre(x_1)' a un cuantificador existencial, obteniendo '$\exists x_1$ Corre(x_1)', en un lenguaje de segundo orden podemos anclar la variable 'X_1' en 'X_1(Susana)' a un cuantificador existencial, obteniendo '$\exists X_1\ X_1$(Susana)'.

El problema de cómo interpretar un lenguaje de segundo orden no es trivial[14]. Pero, como primera aproximación, puede pensarse en

13. Véase, por ejemplo, Drake, 1974.
14. Véase Quine, 1986; Boolos, 1998, especialmente la primera sección, y Rayo y Yablo, 2000.

'$\exists X_1 X_1$(Susana)' como queriendo decir 'existe una propiedad tal que Susana tiene esa propiedad'.

En un lenguaje de segundo orden puede expresarse fácilmente algo parecido a la idea intuitiva de que el esquema de comprensión es válido para *cualquier* propiedad $\ulcorner \mathcal{P}(x) \urcorner$, sea o no expresada por una fórmula del lenguaje:

$$\forall X\, \exists y\, \forall x\, (x \in y \equiv (x \in z \wedge X(x))$$

7. LÓGICAS

Frecuentemente se dice que un lenguaje formal está asociado a una lógica (o, con cierta falta de propiedad, que un lenguaje formal *es* una lógica). Por esto se pueden querer decir dos cosas distintas. En el primer sentido, se dice que un lenguaje formal está asociado a una lógica cuando el lenguaje formal está asociado con un sistema deductivo. En este caso se dice que una fórmula es una verdad lógica cuando es un teorema en el sistema deductivo, y que una fórmula es una consecuencia lógica de un conjunto de fórmulas cuando hay una prueba en el sistema deductivo de la fórmula a partir del conjunto. En el segundo sentido, se dice que un lenguaje formal está asociado a una lógica cuando el lenguaje formal está asociado a una teoría de modelos. En este caso pueden darse caracterizaciones de las nociones de verdad lógica y consecuencia lógica a partir de la teoría de modelos.

En la sección 7.1 veremos qué es una teoría de modelos y cómo puede utilizarse para caracterizar verdad lógica y consecuencia lógica. En la sección 7.2 hablaremos brevemente de las lógicas no clásicas.

7.1. *Teorías de modelos*

Intuitivamente, un enunciado verdadero es *lógicamente* verdadero si su verdad es independiente de cómo es el mundo y del significado de sus partes no lógicas. Intuitivamente, un argumento válido es *lógicamente* válido si el hecho de que sus premisas sean verdaderas sólo si su conclusión es verdadera es independiente de cómo es el mundo y del significado de las partes no lógicas de las premisas y la conclusión.

Una teoría de modelos es una caracterización rigurosa de estas ideas intuitivas. Más específicamente, una teoría de modelos es una clase de 'modelos', cada uno de los cuales representa una manera en que podría ser el mundo y una manera de asignarle significados a las partes no lógicas de los enunciados de un lenguaje. Con base en eso, se define 'verdad lógica' como 'verdadero en todos los modelos', y

'consecuencia lógica de los enunciados en P' como 'verdadero en todos los modelos en los que los miembros de P son verdaderos'.

'Modelo' y 'teoría de modelos' pueden definirse formalmente como sigue. Un *modelo* para un lenguaje formal F es una semántica formal composicional para F (ver sección 5.1). Una *teoría de modelos* T para F es un conjunto de modelos para F con las siguientes características: (*a*) todos los modelos en T seleccionan las mismas secuencias de símbolos básicos como unidades significativas básicas para F; (*b*) existe una función de valores semánticos de unidades significativas básicas a valores semánticos de fórmulas tal que todos los modelos en T le asignan valores semánticos a las fórmulas de F de acuerdo con esa función; y (*c*) existe una función (tal vez parcial) de valores semánticos de fórmulas a valores de verdad tal que todos los modelos en F asignan valores de verdad de acuerdo con esa función.

En la teoría de modelos *clásica* para lenguajes de primer orden los modelos son precisamente aquellos que satisfacen las características de la semántica formal para lenguajes de primer orden de la sección 5.1.

Cuando una unidad significativa básica de F recibe el mismo valor semántico en todos los modelos de T se dice que la unidad significativa es parte del *vocabulario lógico* de F. En las teorías de modelos clásicas para lenguajes de primer orden, el vocabulario lógico consiste de las conectivas proposicionales, el símbolo existencial y el símbolo de identidad.

7.2. *Lógicas no clásicas*

Se dice que la lógica asociada a un lenguaje de primer orden es una *lógica clásica* en dos sentidos distintos. En el primer sentido, esto quiere decir que el lenguaje está asociado con un sistema deductivo clásico (ver sección 3.3). En el segundo sentido, esto quiere decir que el lenguaje está asociado con una teoría de modelos clásica (ver sección 7.1).

Los filósofos no siempre trabajan con lógicas clásicas. En algunos casos esto es porque consideran que las lógicas clásicas no capturan correctamente nuestras nociones intuitivas de verdad y consecuencia lógica. En otros casos es porque consideran que las lógicas no clásicas resultan más útiles para alcanzar ciertos objetivos. Entre las lógicas no clásicas destacan la lógica intuicionista (Troelstra y van Dalen, 1988), la lógica de la relevancia (Read, 1988), las lógicas paraconsistentes (Priest, 1987), las lógicas no monotónicas[15] y las lógicas abductivas (Aliseda, 1997; Flach y Kakas, 2000).

15. Véase, por ejemplo, Antoniou, 1997, y Fariñas del Cerro y Frías Delgado, 1995.

8. CONCLUSIONES

Hemos visto que una formalización es el resultado de parafrasear un enunciado del lenguaje ordinario a un lenguaje formal, más apropiado para el desarrollo de nuestros objetivos. En el primer apartado discutimos las características fundamentales de los lenguajes formales, y en el segundo discutimos la relación entre un enunciado y su formalización. En particular, vimos que, para que nuestras formalizaciones sean correctas, lo único que hace falta es que, a nuestro juicio, nos permitan conseguir algo suficientemente cercano a nuestros objetivos originales.

El tercer y cuarto apartado estuvieron dedicados a los lenguajes de primer orden. Consideramos sus características básicas y los sistemas deductivos con los que suelen estar asociados. En el quinto apartado comparamos tres maneras distintas de interpretar un lenguaje formal: a través de traducciones, a través de semánticas formales y a través de axiomas (o reglas axiomáticas). En el sexto apartado vimos que los lenguajes de primer orden tienen limitaciones expresivas y consideramos dos remedios: los lenguajes modales y los lenguajes de segundo orden. El séptimo apartado estuvo dedicado a las lógicas. Vimos qué es una lógica y en qué consisten las lógicas clásicas.

BIBLIOGRAFÍA

Aliseda, A. (1997), *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*, Institute for Logic, Language, and Computation, Universiteit van Amsterdam, Holland (también disponible en http://www.illc.uva.nl/Publications/Dissertations/DS-1997-04.text.ps.gz).

Antoniou, G. (1997), *Nonmonotonic Reasoning*, MIT Press, Cambridge, MA.

Boolos, G. (1993), *The Logic of Provability*, Cambridge University Press, Cambridge.

Boolos, G. (1998), *Logic, Logic and Logic*, Harvard University Press, Cambridge, MA.

Chellas, B. (1980), *Modal Logic, An Introduction*, Cambridge University Press, Cambridge.

Drake, F. R. (1974), *Set theory: An Introduction to Large Cardinals*, North-Holland Publishing, Amsterdam/London.

Fariñas del Cerro, L. y Frías Delgado, A. (1995), «Razonamiento no monótono: Un breve panorama»: *Theoría*, S. E., vol. X/23, 7-26.

Flach, P. y Kakas, A. (eds.) (2000), *Abductive and Inductive Reasoning: Essays on their Relation and Integration*, Kluwer Academic Publishers, Applied Logic Series.

Gentzen, G. (1969), *The Collected Papers of Gerhard Gentzen*, en M. E. Szabo (ed. y trad.), North-Holland Publishing, Amsterdam/London.

Hintika, J. e Hintika, M. B. (1989), *The Logic of Epistemology and the Epistemology of Logic*, Kluwer Academic Publishers, Dordrecht.
Koslow, A. (1992), *A Structuralist Theory of Logic*, Cambridge University Press, Cambridge.
Lewis, D. (1986), *On the Plurality of Worlds*, Blackwell, Oxford.
McGee, V. (1999), «Everything», en G. Sher y R. Tieszen, *Between Logic and Intuition: Essays in Honor of Charles Parsons*, Cambridge University Press, Cambridge.
Mendelson, E. (⁴1997), *Introduction to Mathematical Logic*, Chapman & Hall, London.
Orayen, R. (1989), *Lógica. Significado y Ontología*, UNAM, México.
Priest, G. (1987), *In Contradiction*, Nijhoff, Dordrecht.
Prior A. (1961), «The Runaway Inference Ticket»: *Analysis*, 21, 129-131.
Quine, W. V. (1951), *Mathematical Logic*, Harvard University Press, Cambridge, MA.
Quine, W. V. (1960), *Word and Object*, MIT Press, Cambridge, MA.
Quine, W. V. (²1986), *Philosophy of Logic*, Harvard University Press, Cambridge, MA.
Rayo, A. y Yablo S. (2000), «Nominalism Through De-Nominalization»: *Noûs*, 35/1.
Read, S. (1988), *Relevant Logic*, Blackwell, Oxford.
Tarski, A (1933), «The Concept of Truth in Formalized Languages», en Íd., *Logic, Semantics and Metamathematics*, Hacket, Indianapolis.
Troelstra, A. S. y van Dalen, D. (1988), *Constructivism in Mathematics: An Introduction*, 2 vols., North-Holland Publishing, Amsterdam/London.

LÓGICA Y PSICOLOGÍA*

Gladys Palau

1. INTRODUCCIÓN

Es posible abordar la problemática en torno a las relaciones entre lógica y psicología o bien desde una perspectiva sistemática, o bien desde una perspectiva histórica. Sin restar valor a la primera, en este trabajo nos hemos inclinado por la perspectiva histórica porque somos de la opinión de que la elucidación de ciertos problemas filosóficos —tales como los que aquí trataremos— se enriquece sustancialmente cuando se indaga en los debates filosóficos en los que históricamente se han constituido. Por esta razón y dada la amplitud de la problemática epistemológica involucrada, nos restringiremos a exponer sólo aquellos enfoques teóricos que en la historia de la filosofía, y particularmente de la lógica, han producido cambios sustanciales en la forma de abordar tanto la naturaleza de la lógica como las relaciones entre la lógica y la psicología.

Es sabido que la lógica y la psicología nacieron de la filosofía. Sin embargo su desarrollo posterior se vio condicionado por ciencias de muy distinta naturaleza: la lógica por las ciencias formales y la psicología por las ciencias empíricas. Es sabido también que las ciencias formales (llamadas comúnmente *demostrativas* o *deductivas*) fueron las primeras en constituirse históricamente, al punto de afirmarse que la lógica, en tanto ciencia formal, fue creada por Aristóteles *ex nihilo*. En cierto sentido esta afirmación es correcta, ya que su teoría sobre el silogismo y su caracterización de ciencia demostrativa constitu-

* Al final del presente trabajo se incluye un gráfico sobre el panorama histórico del debate sobre el psicologismo.

yeron las primeras reflexiones teórico-sistemáticas sobre el objeto de la lógica y el concepto de demostración, que a su vez posibilitaron la constitución posterior de la geometría como ciencia en *Los Elementos* de Euclides. Por el contrario, aunque gestadas en el Renacimiento, las ciencias naturales y experimentales sólo se constituyen como tales en el siglo XVII. En efecto, aun cuando el término *psicología* fue introducido por primera vez en el año 1590 por Goclenius como designando a la ciencia del alma, en tanto forma y principio del cuerpo, la psicología como ciencia empírica surge muy posteriormente, más precisamente entre los años 1860-1880 y bajo la nominación de *nueva psicología* o *psicología experimental*, vía los trabajos de los pensadores alemanes Stumpf, Müller, Brentano, Ebbinghaus, Külpe y Wundt. Si bien somos de la opinión de que el estudio sobre las relaciones entre la lógica y la psicología se instala abiertamente en el debate sobre el llamado *psicologismo* acaecido dentro de la filosofía académica alemana entre los años 1880 y 1920, sostenemos que el germen de ese debate se encuentra en el enfoque psicológico de la lógica plasmado en *La Logique ou l'art de penser* de A. Arnauld y P. Nicole, conocida luego como *Lógica de Port Royal,* aparecida en 1662 y de vigencia continua hasta 1870, cuando la psicología se convierte en ciencia experimental. Por ello comenzaremos nuestra exposición con un esbozo de las posiciones que consideramos más representativas de este período, con el objeto de mostrar, por un lado, que en ellas se gestan las ideas directrices defendidas por los psicologistas posteriores, y por el otro, que en ninguna de ellas se encuentra una clara distinción entre el objeto de la lógica y el objeto de la psicología, distinción que nosotros consideramos condición necesaria para entender en profundidad la controversia sobre el psicologismo.

2. EL ENFOQUE PSICOLÓGICO DE LA LÓGICA

Uno de los primeros exponentes del enfoque psicológico de la lógica fue J. Jungius, quien, en su obra *Logica Hamburgensis* de 1635, sostiene que la lógica es el arte de dirigir las operaciones de la mente humana tendientes a distinguir lo verdadero de lo falso, a saber: el concepto o noción, la sentencia o juicio y el discurso. Aunque algunos comentaristas consideren a la lógica de Jungius más importante y rica que la *Lógica de Port Royal*, ésta es la que posteriormente se convirtió en referencia indiscutida de este período de las investigaciones acerca de la naturaleza de la lógica. Entroncada en la tradición agustiniana, vía Descartes, la *Lógica de Port Royal* no comparte el antiaristo-

telismo radical de Petrus Ramus, sostenido en su obra *Dialectique* de 1555, sino que más bien toma una posición crítica frente a la tradición lógica escolástica y se caracteriza por adoptar un lenguaje psicológico para la descripción de los objetos que se considera deben ser estudiados por la lógica. En efecto, en el Discurso II de la *Logique ou l'art de penser*, al defender el nombre puesto a la obra, sus autores afirman:

> [...] puesto que el fin de la lógica es reglar para todas las operaciones de la mente, tanto para las ideas simples como para los juicios y razonamientos, raramente haya otra palabra que incluya todas estas operaciones: y la palabra *pensamiento* ciertamente comprende a todas ellas, porque las ideas simples son pensamientos, los juicios son pensamientos y los razonamientos son pensamientos (1851, 14).

En las tres primeras partes la obra se ocupa de los tres temas lógicos tradicionales, o sea, del concepto (ideas simples), de los juicios y de los razonamientos, y a los efectos de poder definir a la lógica como el arte de dirigir la razón acertadamente, agregan el método como cuarto tema. Por ello, para los autores, la tarea de la lógica consiste en reflexionar sobre las cuatro operaciones principales de la mente: concebir (ideas), juzgar, razonar y ordenar. Sin embargo, para ellos, la lógica no consiste en descubrir el modo como se forman estas operaciones, ya que ellas se constituyen naturalmente, sino en reflexionar acerca de lo que la naturaleza hace «dentro de nosotros». Por ejemplo, debemos reflexionar sobre la naturaleza y origen de las ideas, sobre su simplicidad o complejidad, sobre su individualidad y universalidad, sobre su claridad, oscuridad, distinción y confusión. Coincidimos con A. Coffa (1991, 9) cuando afirma que la *Lógica de Port Royal* canoniza la tradición según la cual las ideas o representaciones son el objeto de la lógica, puesto que ellas son las que nos posibilitan conocer lo que está «fuera de nosotros». La influencia cartesiana es, pues, evidente y, al decir de Benson Mates (1974), la forma de tratar los temas lógicos nada tiene que ver con lo que hoy en día se entiende por lógica y justifica que algunos autores consideren a esta obra uno de los primeros claros exponentes del psicologismo en lógica. Pese a que ya existían en ese momento indicios de otra forma de pensar la lógica, tal como lo veremos en el parágrafo 4, la *Lógica de Port Royal* predominó en la concepción de la lógica hasta finales del siglo XVIII en pensadores franceses como E. B. de Condillac en sus obras *Logique* de 1780 y *Langue des calculs* de 1798 y en ingleses como H. Aldrich en *Artis Logicae Compendium* de 1691.

El enfoque psicológico de la lógica recibe un segundo impulso del movimiento filosófico alemán de comienzos del siglo XIX, conocido con el nombre de *naturalismo* o *psicologismo filosófico*. Si bien es cierto que el objetivo de este movimiento no fue analizar la naturaleza de la lógica, hemos decidido su inclusión porque tanto Husserl como Frege hacen referencia a integrantes de este movimiento en su polémica contra el psicologismo y porque en la literatura es común que se presente a los integrantes de este movimiento como los últimos exponentes de la filosofía moderna post-kantiana opuesta al idealismo de Fichte, Schelling y Hegel y, en tanto tales, proclives a aceptar un enfoque psicológico de la lógica.

El primer representante del naturalismo filosófico fue Jakob Fries (1773-1843), profesor en Heildelberg desde 1805 y enemigo de Hegel también en lo personal. Enmarcado dentro de un subjetivismo ético y religioso, propuso una revisión de la teoría de la razón pura de Kant, y en su *Neue Kritik der Vernunft* de 1803 (revisada posteriormente en 1938 bajo el nombre de *Anthropologische Kritik der Vernunft*) sostiene que los principios de la razón son efectivamente *a priori* pero que sin embargo pueden ser conocidos «antropológicamente», *i.e.*, por observación de nuestras facultades cognoscitivas. De ahí que para Fries la actividad filosófica sea considerada como una especie de reflexión sobre las experiencias internas susceptible de ser realizada mediante un análisis psicológico (o antropológico). De su escuela surge Friedrich Beneke (1798-1854), quien, en sus obras *Lehrbuch der Logik als Kunstlehre des Denkens* de 1832 y *System der Logik als Kunstlehre des Denkes* de 1842, consideró también a la psicología como la ciencia filosófica fundamental para realizar el análisis de la experiencia interna. En su intento de reemplazar la filosofía especulativa por una ciencia que respondiera al modelo de la ciencia natural, Beneke postuló a la psicología como la base de todas las ciencias y en particular de la filosofía. Sin embargo, Beneke rechazó considerar a la psicología como doctrina de las facultades mentales y conceptos de carácter innato y se inclinó por sostener que ella debía dedicarse a la investigación de la facultad del entendimiento (*Verstand*), es decir, al origen y a las formas del pensamiento. A diferencia de Fries, Beneke se ocupó particularmente de la relación entre lógica y psicología, afirmando que la psicología es la base fundacional de la lógica.

Como tercera contribución al enfoque psicológico de la lógica en este período hemos decidido incluir los aportes de John Stuart Mill (1806-1873). Contemporáneo del naturalismo filosófico y surgido de una tradición filosófica profundamente distinta, la obra de Stuart Mill

ha sido interpretada de diversas maneras: E. W. Beth (1968) lo considera una variante de psicologismo; W. y M. Kneale (1972) lo ubican en una posición ambivalente, rechazando el convencionalismo de Hobbes y el psicologismo del naturalismo alemán y otros ni siquiera lo mencionan en la historia de la lógica (Bochenski, 1976, y Mates, 1974). Cualquiera sea la posición que se asuma, es innegable que, tal como lo afirma M. Kusch, en su excelente y exhaustivo estudio titulado *Psychologism* (1995), las ideas de Stuart Mill respecto de la naturaleza de la lógica y la matemática, inducen a una visión psicológica de las mismas que justifica plenamente su inclusión en este período sobre la naturaleza de la ciencia de la lógica. Más aún, es conocida la crítica que Frege hace a la filosofía de la matemática de Mill en sus *Grundlagen der Arithmetik*.

Ateniéndonos lo más extrictamente posible a lo expuesto por Stuart Mill en *System of Logic* —escrito en 1843—, el objetivo esencial de esta obra es brindar una exposición de la lógica coherente con el empirismo. En su Introducción, titulada *Definition and Province of Logic*, Mill brinda una descripción bastante vaga acerca del objeto de la lógica, a saber:

> La Lógica ha sido definida como arte y como la ciencia del razonamiento; entendiendo por lo primero el análisis de los procesos mentales que tienen lugar cuando se razona y por lo segundo las reglas, basadas en ese análisis, cuyo fin es dirigir correctamente tal proceso (1967, 1-2).

Luego, dedica varios capítulos al análisis de la inferencia o razonamiento de la lógica formal, sin realizar referencia alguna a procesos mentales y siempre en relación con la teoría del silogismo. Acepta también la versión tradicional del silogismo según la cual un silogismo es un razonamiento (deductivo) que va de lo general a lo particular, pero se pregunta: ¿es realmente el silogismo un proceso de inferencia que va de lo conocido a lo desconocido? Y, luego de analizar los modos silogísticos válidos, afirma que la respuesta a la pregunta planteada debe ser negativa, ya que en todos los casos la conclusión de cualquier silogismo está presupuesta en las premisas, constituyendo por lo tanto una *petitio principii* que no sirve para probar nada más que lo que ya está dicho en los principios de partida. Por ello, para Stuart Mill, las inferencias de la lógica deductiva no son auténticas inferencias. La inferencia real (auténtica) reside en el camino que permite llegar a la verdad de la premisa general, es decir, en cómo se llega a las verdades generales a partir de las verdades particulares. De

ahí que, desde su punto de vista, el auténtico razonamiento debe ser identificado con la *inducción*. Por ello, según Stuart Mill, la incumbencia de la lógica debe restringirse a esa porción del conocimiento que consiste en inferencias a partir de verdades previamente conocidas y la lógica es por lo tanto la *ciencia de la prueba*, o *evidencia*, o, como él mismo la llama, *Lógica de la Verdad* (1967, 5). A la pregunta sobre cuál es la evidencia sobre la que reposan los axiomas de la Lógica de la Verdad, Mill responde: ellos son verdades experimentales, generalizaciones a partir de la observación), o sea, generalizaciones empíricas (1967, 151). Consecuentemente, las verdades necesarias de la lógica formal (como por ejemplo el *Principio de no-contradicción* y el *Tercero Excluido*), aun cuando sus negaciones sean inconcebibles, son también para Mill simples generalizaciones de experiencias mentales universales y el objeto de la lógica en tanto arte es precisamente el estudio de esas experiencias mentales universales.

Por otra parte, pese a considerar a la psicología una ciencia genuina, en *System of Logic* no aparece afirmación alguna en la cual Mill sustente que, a pesar de que ambas disciplinas tengan el mismo carácter empírico y su método sea la inducción, la psicología (en tanto ciencia genuina e independiente) sea la ciencia fundacional de la lógica (en tanto arte). Su posición es más radical y afirma categóricamente que la lógica en tanto arte es un brazo de la psicología, ya que la psicología utiliza la inducción como método para descubrir verdades y la lógica (en tanto arte) también utiliza la inducción como método para descubrir las leyes universales del pensamiento. Así, distinguir en la obra de Mill entre lógica (en tanto arte) y psicología constituye una dificultad evidente.

3. EL PSICOLOGISMO EN LÓGICA

El psicologismo en lógica, en tanto corriente filosófica explícita, se constituye simultáneamente, como ya se dijo, con la aparición de la psicología experimental, en particular con la llamada *psicología fisiológica* de Wilhelm Wundt. Pensadores como Ebbinghaus, Külpe, Stumpf y aun Brentano y Meinnong contribuyeron o se adhirieron a esta *nueva psicología*. Pero no todos ellos se ocuparon de la relación entre lógica y psicología. Entre los que se ocuparon de esta problemática, se encuentran el mismo Wilhelm Wundt (1853-1920), Christopher Sigwart (1830-1904), Theodor Lipps (1851-1914) y Benno Erdmann (1855-1921), los tres primeros comúnmente considerados como los representantes máximos del psicologismo alemán del siglo XIX.

Nuestra opinión coincide con Martin Kusch, quien en su ya mencionado libro sitúa la controversia sobre el psicologismo entre los años 1870 y 1920. Como dato interesante, menciona además que el termino *psicologismo* (*psychologismus*) fue usado por primera vez por el hegeliano Eduard Erdmann en 1886 para referirse a la filosofía naturalista de Beneke y que se popularizó después de las críticas que recibió de parte de Husserl. Por nuestra parte, asumiendo la imposibilidad de argumentar en este trabajo a favor o en contra de las distintas acepciones que posteriormente se diera al término psicologismo, convenimos en llamar *psicologismo* al enfoque de la lógica que, habiendo aceptado a la psicología como genuina ciencia natural, reduce, de una u otra forma, la explicación de las nociones lógicas y leyes lógicas a procesos psicológicos y leyes psicológicos respectivamente.

Cronológicamente, Christopher Sigwart fue el primero en destacar, por un lado, la necesidad de crear una doctrina del arte de pensar que sirva de guía para un análisis lógico del pensamiento a fin de «arribar a juicios seguros y válidos universalmente», y por el otro, la imposibilidad de realizar este objetivo sin un análisis de los actos mentales involucrados. En su obra *Logik* (1873) define a la lógica como el estudio de las normas y métodos del pensar, convirtiéndose así en el primer exponente del psicologismo y por ello en uno de los blancos principales de los posteriores ataques de Husserl.

Sin embargo, fue Wilhelm Wundt, dada su contribución a la constitución de la nueva psicología experimental, quien posibilitó la construcción de un nuevo marco conceptual que permitiera a la psicología, primero, cumplir con los requisitos de una ciencia natural; segundo, independizarse definitivamente de la metafísica y terminar con el dualismo cartesiano y la introspección «pura» como objeto de la psicología; tercero, independizarse como ciencia genuina diferente también de la fisiología; y cuarto, caracterizar claramente el objeto de la psicología como el estudio de los fenómenos elementales de la experiencia interna mediante la «introspección experimental». Su voluminosa obra *Logik*, de 1883, comienza precisamente considerando a la lógica como la ciencia que debe dar cuenta de las leyes que operan en el conocimiento científico, ya que ella es posible gracias a la existencia de este tipo de conocimiento. Así, el objeto de la lógica consiste en «apartar» en los diversos enlaces de representaciones que componen el conocimiento aquellos que proporcionan leyes para el desenvolvimiento del saber. En otras palabras, la lógica es una ciencia normativa que debe identificar aquellas combinaciones de ideas que son universalmente válidas. En síntesis, mientras la psicología se ocupa de averiguar cómo los seres humanos piensan, la lógica debe ocupar-

se de cómo se debe pensar a fin de lograr el conocimiento científico. Dado que las leyes psicológicas carecen de las propiedades de evidencia y universalidad propias de las leyes lógicas, ellas son perfectamente diferenciables de estas últimas. Sin embargo, a ambos tipos de leyes se llega por generalización y observación de pensamientos empíricos y por ello la explicación y formulación de las leyes lógicas siempre e inevitablemente contienen, según Wundt, conceptos psicológicos. En otros términos, puesto que el conocer se da sólo en la psiquis, lo que en la lógica se piensa es un hecho psicológico y, por ello, los objetos de estudio de la lógica deben explicarse mediante conceptos de la nueva psicología y la lógica es una disciplina psicológica.

Desde la filosofía alemana Theodor Lipps fue quien asumió la defensa más explícita y clara de la corriente psicologista, convirtiéndose así en el representante de mayor influencia filosófica posterior. En su obra *Grundzüge der Logic*, de 1893, sostiene que la lógica es la ciencia de las formas y leyes del pensamiento y que se la puede definir como la actividad del espíritu por la cual «lo dado a la conciencia se convierte en razonamiento». Coincidiendo con Wundt, afirma que las leyes de la lógica son leyes naturales del pensamiento y que la lógica, dado su carácter normativo, es la «física» del pensamiento.

Sin embargo, el psicologismo no fue aceptado en forma unánime dentro de la academia filosófica alemana. La tradición idealista en la filosofía alemana había sido demasiado fuerte como para aceptar sin más la naturalización de la filosofía y la reducción de la lógica a una rama de la ciencia experimental. La crítica al psicologismo no se hizo esperar y se encarnó en dos filósofos también alemanes: Edmund Husserl (1859-1938), discípulo de Brentano y formado en matemática y astronomía en la Universidad de Leipzig, y Gottlob Frege (1848-1925), filósofo y matemático, fundador de la lógica matemática contemporánea y de influencia determinante en determinadas corrientes de la filosofía contemporánea como el empirismo lógico y la filosofía analítica.

4. LA REFUTACIÓN DEL PSICOLOGISMO DE HUSSERL Y FREGE

En *Philosophie der Arithmetik*, de 1891, y bajo la influencia de Brentano, al preguntarse sobre el origen de los conceptos y de las intelecciones matemáticas, Husserl elabora una respuesta basándose en la distinción de Brentano entre representaciones propias e impropias (simbólicas), a la cual responde afirmando que el conocimiento matemático se funda en representaciones simbólicas, adoptando de este

modo una posición psicologista ante la matemática. Sin embargo, esta posición será prontamente abandonada por Husserl. En efecto, en 1894 Frege publica un trabajo criticando la filosofía de la aritmética de Husserl, publicado en *Zeitschrift für Philosophie und Philosophische Kritik*, y en el cual muestra que el tratamiento de Husserl respecto de la naturaleza psicológica de los números conduce a consecuencias inaceptables. Husserl acusa el impacto de la crítica y en sus *Investigaciones lógicas* (*Logische Untersuchungen*), escrito entre 1900-1901, después de confesar la tendencia psicologista de su primera etapa de pensamiento, abandona definitivamente el psicologismo. Consecuentemente, en los Prolegómenos de la misma obra, se dedicará a refutar toda forma de psicologismo, en particular el de los filósofos naturalistas Beneke, Sigwart y Lipps. En los dos primeros capítulos Husserl analiza los problemas teoréticos que plantean las disciplinas particulares y define el sentido por el cual la lógica (o teoría de la ciencia) puede ser considerada una disciplina normativa y práctica (o arte). Los capítulos siguientes los destina a criticar al psicologismo tanto respecto de sus argumentos como de sus consecuencias filosóficas, y a defender la tesis según la cual los fundamentos teóricos de la lógica, en tanto disciplina práctico-normativa, no son ni psicológicos ni biológicos. Sus críticas al psicologismo son de dos tipos. Las correspondientes al primer tipo son presentadas bajo la misma forma indirecta de argumentación: si se aceptan las tesis psicologistas se obtienen determinadas consecuencias empíricas que pueden ser refutadas; luego el psicologismo es insostenible. Las críticas del segundo tipo consisten en mostrar que las tesis psicologistas involucran prejuicios cuya falsedad es fácilmente demostrable. A nuestro propósito nos interesa destacar en forma sintética y simple solamente las consecuencias que Husserl infiere a partir de la tesis psicologista según la cual las leyes de la lógica son leyes psicológicas o se fundan en leyes psicológicas. Tales consecuencias serían, entre otras, las siguientes: C1) las leyes de la lógica serían tan vagas e imprecisas como las de la psicología; C2) las leyes de la lógica, por igual motivo que en C1, no serían leyes *a priori*; C3) las leyes de la lógica harían referencia a entidades psicológicas; y C4) las leyes de la lógica estarían relativizadas a cada especie humana, ya que toda forma de psicologismo conduce al relativismo. Pero puesto que: H1) las leyes de la lógica no son vagas, sino exactas y precisas; H2) son evidentes, apodícticas y *a priori*; H3) no involucran entidades psicológicas, y H4) las leyes de la lógica son invariantes y no tienen excepciones, se sigue que las leyes de la lógica no son leyes psicológicas. A partir de tales críticas, Husserl concluye que el psicologismo ha sido definitivamente refutado (Kusch, 1995, 43-58).

Sin embargo, aún le queda a Husserl sin responder la pregunta acerca del fundamento de la lógica, ya que todo conocimiento científico es un conocimiento *por fundamentos* (1967, 260). Pero como además las ciencias normativas dependen de las teoréticas (en el sentido de nomológicas) (1967, 265) y la lógica normativa es la lógica de la ciencia, ella debe fundarse en una disciplina teorética, la cual será bautizada por él *lógica pura*. Pero ¿qué entiende Husserl por *lógica pura*?

En el capítulo XI de los *Prolegómenos*, titulado «La Idea de la posibilidad de una teoría general de la ciencia (o de la unidad deductiva en general)» Husserl afirma: «[...] la lógica pura abarca del modo más universal las condiciones ideales de la posibilidad de la ciencia en general» (1967, 284). En otras palabras, la lógica pura consiste en describir las leyes que expresan las condiciones ideales de la posibilidad del conocimiento deductivo y teorético en general. En síntesis, para Husserl, la lógica pura se ocupa del concepto (unidad ideal de significación), del juicio (situación objetiva producto del acto de juzgar) y del raciocinio (contenidos unitarios del acto de razonar), todas ellas en tanto unidades ideales o significaciones, es decir, aisladas de las expresiones gramaticales y de las diferentes vivencias en las que son pensadas.

La primacía del enfoque psicológico de la lógica que se iniciara con la *Lógica de Port Royal* hizo que la tradición Leibniz-Bolzano permaneciera casi ignorada. En efecto, pese a que las obras de juventud de Leibniz, particularmente *Ars combinatoria*, escrita en 1664, es contemporánea de *La Logique de Port Royal*, su *Calculus ratiocinator* no se publicará hasta el siglo XIX. En *Ars combinatoria*, al sostener la posibilidad de construir un lenguaje simbólico artificial cuya estructura fuese reflejo del pensamiento y que permitiera liberar al estudio lógico de las vaguedades del lenguaje ordinario, Leibniz inicia una forma de encarar la problemática lógica desde el paradigma del pensamiento matemático. Su formación filosófica, enrolada en el más puro racionalismo, y sus contribuciones a la matemática que lo sitúan como precursor del logicismo contemporáneo, condicionaron una concepción de la lógica independiente de todo matiz psicológico, tal como se comprueba en su teoría de la identidad, el tratamiento de la noción de necesidad y la consecuente caracterización de las verdades lógicas como verdades de razón o verdades en todos los mundos posibles.

El enfoque de la lógica iniciado por Leibniz recibirá posteriormente contribuciones menores como la de Christian Wolff en su *Philosophia rationalis* de 1728 y otras decisivas como la de B. Bolzano (1781-1848), en cuyas obras *Wissenschaftslehre* de 1837 y *Paradoxien des Unendlichen* de 1851 expone ideas originales sobre lógica, que marcarán el pensamiento lógico posterior, como, por ejemplo, que una

proposición es un objeto real cuya verdad o falsedad es independiente de cualquier contexto, *i.e.*, son independientes del sujeto que las piensa; que el concepto de verdad es objetivo, *i.e.*, no es ni epistémico ni psicológico; y que una proposición es universalmente válida cuando todas sus «variantes» son verdaderas.

Sin embargo, la historia de las ciencias muestra que el cambio de perspectiva filosófica en el análisis de determinados problemas muchas veces se genera o resurge de logros cognoscitivos provenientes de otras disciplinas afines. Tal es lo que sostenemos ha sucedido con la tradición logicista iniciada por Leibniz. En efecto, la constitución de la lógica matemática hacia fines del siglo XIX aparece estrechamente vinculada al desarrollo del álgebra y a la fundamentación de la matemática. Desde sus orígenes y hasta el siglo XIX el dominio del álgebra se había reducido a la búsqueda de métodos de resolución de ecuaciones cada vez más generales. En 1799 Gauss demuestra en su tesis doctoral el hoy conocido como *Teorema fundamental del álgebra*; pocos años después Galois logra las primeras estructuras de grupo y en 1879 Dedekind define por primera vez la estructura de cuerpo. Había nacido una nueva matemática pura que a su vez involucraba una nueva concepción de la demostración matemática. En este contexto surgen las obras de los algebristas ingleses George Boole (1815-1864) y Augustus De Morgan (1806-1871), incluidos sin discusión entre los iniciadores de la lógica matemática contemporánea. En 1847 aparece la obra *Mathematical Analysis of Logic*, de Boole, en la cual esboza su idea básica según la cual las fórmulas algebraicas pueden perfectamente ser usadas para expresar relaciones lógicas, tal como lo muestra el cálculo lógico de clases, construido mediante la aplicación de los procedimientos formales del álgebra simbólica y conocido después como *álgebra de la lógica*. Es interesante hacer notar que en 1854 escribe *An Investigation of the Laws of Thought on which are founded the Mathematical Theory of Logic and Probability* y en la cual Boole intenta realizar una distinción entre la lógica de clases y una especie de lógica de nivel superior, en el sentido de una filosofía capaz de expresar en símbolos cualquier pensamiento (Kneale y Kneale, 1972). Simultáneamente a la aparición de *Mathematical Analysis of Logic*, de Boole, en 1847, Augustus De Morgan, también algebrista de profesión, publica su obra *Formal Logic.* Su aporte principal consistió en presentar una versión algebraica de la teoría del silogismo en armonía completa con las ideas expuestas por Boole, y más tarde como un caso particular de la teoría de las relaciones. Posteriormente, el moderno enfoque de la lógica de Boole y De Morgan trasciende a los Estados Unidos, donde recibe importantes aportes de C. S. Peirce, y lue-

go a Alemania, donde será desarrollado plenamente por Frege, quien precisamente tratará de aplicar los métodos de la nueva lógica simbólica a la fundamentación de la matemática. En su *opera prima* de 1879, titulada *Begriffsschrift* (*Notación conceptual*), Frege brinda la primera presentación axiomática de la lógica de primer orden. En 1884 publica *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchugun ubre den Begriff der Zahl* (*Fundamentos de la aritmética: una investigación lógico-matemática sobre el concepto de número)* y en 1893-1903, *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet* (*Leyes básicas de la aritmética: una exposición del sistema).* En ellas no sólo expone su programa logicista acerca de la matemática, consistente en presentar a la matemática como una extensión de la lógica, sino que también ofrece las primeras críticas a las formas vigentes en esa época de encarar la naturaleza de la matemática y la lógica, haciendo referencia específica a la lógica psicológica de Fries y Beneke. Sus críticas al psicologismo se completarán varios años después, en su artículo *Der Gedanke: eine logische Untersuchung (Pensamientos),* escrito entre 1918 y 1923.

Kusch (1995, 32-38) expone detalladamente la posición de Frege frente a la relación entre lógica y psicología sostenida en *Die Grundlagen* y en *Grundgesetze* mediante un conjunto de tesis, entre las cuales hemos seleccionado aquellas que consideramos las más representativas a nuestro propósito, a saber: F1) la matemática y la lógica no son parte de la psicología y la verdad de sus leyes no es probada por observaciones psicológicas o leyes psicológicas) (*Die Grundlagen*); F2) mientras la matemática y la lógica son ciencias precisas y exactas, la psicología es vaga e imprecisa, por lo cual no es plausible que la matemática y la lógica sean parte de la psicología (*Die Grundlagen*); F3) la distinción entre juicios analíticos y sintéticos, *a priori* y *a posteriori*, se basa únicamente en la forma en que su verdad es justificada (*Die Grundlagen*); F4) las leyes lógicas no están indexadas a ninguna especie, es decir, son tópicamente neutras (*Grundgesetze)*; F5) la palabra *idea* no debe usarse nunca para hacer referencia a objetos y conceptos, sino solamente para referirse a entidades psicológicas (*Die Grundlagen*), y F6) la lógica y la psicología son ciencias independientes porque la psicología se ocupa de lo «subjetivo» mientras que la lógica debe ocuparse de lo «objetivo», *i.e.*, la lógica es independiente de la conciencia, de las sensaciones y de la experiencia interna (*Der Gedanke*).

Para los fines del presente trabajo, la tesis F4) merece un comentario especial, ya que su fundamentación involucra un análisis más detallado de las relaciones entre las leyes lógicas y las leyes psicológicas.

En efecto, según Kusch (1995, 33-38), en el prólogo a *Grundgesetze* Frege distingue entre leyes descriptivas (como las leyes de la física) y leyes presciptivas (como las de la moral), y acusa a los partidarios de la lógica psicológica, especialmente a B. Erdmann, de identificarlas. La argumentación de Frege parece ser la siguiente: 1) Toda ley descriptiva puede ser aprendida o reformulada como una prescripción para pensar según ella y, así, conlleva una ley prescriptiva, es decir, una ley del pensamiento, ya que ellas legislan sobre cómo se debe pensar. 2) Las leyes lógicas descriptivas pueden ser comprendidas y reformuladas como prescriptivas, es decir, como leyes del pensamiento, porque ellas legislan todo pensamiento y no son tópico-dependientes. Luego, el gran error del psicologismo es creer que las leyes del pensamiento son leyes psicológicas descriptivas o que son leyes prescriptivas que derivan de leyes psicológicas descriptivas. Si así lo fueran, las leyes lógicas serían relativas a cada especie humana y, puesto que ni siquiera las leyes lógicas descriptivas lo son, entonces las leyes de la lógica son independientes de las leyes de la psicología. Por análogas razones, la lógica como ciencia no puede responder si sus leyes descriptivas básicas son verdaderas o no. Evidentemente, no parece quedar lugar en el pensamiento de Frege para una reducción de la lógica a la psicología, pero sin embargo el supuesto de existencia de leyes lógicas descriptivas y su paso a las leyes prescriptivas deja abierto un camino para hipotetizar acerca de una posible relación entre ambas ciencias y que Frege no llega a elucidar claramente. Retomaremos este problema al final del presente parágrafo.

De acuerdo a las tesis recién expuestas, hay buenas razones para afirmar que Frege y Husserl coinciden en el núcleo de la argumentación en contra del psicologismo: 1) que las leyes de la lógica no son leyes psicológicas, porque si lo fueran, ellas serían vagas e imprecisas y no serían conocidas *a priori* (H1, H2, F1, F2); 2) que las leyes de la lógica son universales y no son relativas a ninguna especie (H4, F4); y 3) que los objetos de la lógica (y de la matemática) no son entidades psicológicas (H3, F5, F6). Sin embargo, no todos sus comentaristas están de acuerdo en el rol de ambos en el debate contra el psicologismo.

En la presentación tradicional Husserl es considerado como el autor que refutó definitivamente al psicologismo. Otros, como Føllesdal (1958), sostienen que los argumentos de Husserl no son otra cosa que una adaptación de las críticas de Frege. Otros van todavía más allá y, como E. W. Beth (1968, 44), basándose en determinadas argumentaciones husserlianas, presentadas en textos muy diferentes, sostienen que el antipsicologismo de Husserl no está dirigido a la constitución de una nueva lógica (prueba de ello es que no ha realizado contribu-

ción alguna a dicha ciencia), sino que tiene por fin la necesidad de fundamentar una psicología descriptiva o fenomenológica, análoga a la física y a la química fenomenológicas, tal como las propagaban Duhem y Mach, y que por lo tanto en la obra de Husserl no se encuentra una ruptura radical con el psicologismo. Nosotros sostenemos al respecto una posición conciliadora en el sentido de que, por un lado, creemos que Husserl concibe una lógica independiente de la psicología, cuyos objetos de estudio son entidades ideales, pero por el otro sostenemos que no logra desprenderlos totalmente de cierta carga psicológica. En efecto, cuando afirma: «la lógica pura se ocupa del concepto (unidad ideal de significación), del juicio (situación objetiva producto del acto de juzgar) y del raciocinio (contenidos unitarios del acto de razonar), en tanto unidades ideales o significaciones aisladas de las expresiones gramaticales y de las diferentes vivencias en las que son pensadas y a las que se llega por medio de análisis fenomenológico» (1967, 292-293), aparece como diferenciando claramente la lógica de la psicología. Pero cuando en el capítulo IV de la *Investigación Primera*, a propósito del carácter no consciente de la significación, dice: «No podemos pronunciar nunca la palabra 'luego', sin reportar la mirada sobre el contenido de la significación que tienen las premisas. Al juzgar las premisas, no sólo vivimos en los juicios, sino que reflexionamos sobre los contenidos de ellos; sólo con referencia a esos contenidos aparece motivada la conclusión», no podemos dejar de encontrar resabios del pensar psicológico acerca de la lógica.

Por el contrario, creemos que la posición de Frege es más radical al respecto. En efecto, en *Der Gedanke* afirma:

> Para evitar cualquier mala interpretación e impedir que se borre la frontera entre la lógica y la psicología, asigno a la lógica la tarea de encontrar las leyes del ser verdad, no las del tener por verdadero o las del pensar.

Pero podría objetarse que, en *Der Gedanke,* lo que es verdad son los «pensamientos», el ser de un pensamiento es su verdad y el fin de toda ciencia es descubrir pensamientos verdaderos, y por ello también hay en Frege resabios del enfoque psicológico. Sin embargo, creemos que el uso de la palabra *pensamiento* no constituye en Frege una concesión a la psicología, ya que ella tiene un significado no psicológico, pues los pensamientos no son concebidos como entidades mentales, sino en tanto informaciones o «significados» de oraciones aseverativas. En otras palabras, como afirma Susan Haack (1979, 240), el sentido de una oración no es para Frege una entidad mental o idea, sino un pensamiento en tanto proposición u objeto abstrac-

to. Así, uno de los objetivos de la lógica creada por Frege será establecer las leyes de cómo componer pensamientos verdaderos (o proposiciones) a partir de otros pensamientos (o proposiciones) también verdaderos.

Por nuestra parte, somos de la opinión de que la mayor prueba del antipsicologismo de Frege y de su ruptura filosófica con toda forma de psicologismo reside en sus obras propiamente lógicas y en su programa logicista respecto de la matemática y de que, pese a la envergadura de sus contibuciones lógicas, en su pensamiento resta aún un lugar para establecer una relación entre la lógica y la psicología. Nótese que en *Grundgesetze* Frege había afirmado que la lógica, en tanto ciencia, no podía responder por qué son verdaderas ciertas leyes lógicas descriptivas básicas, que las leyes lógicas prescriptivas se justifican en referencia a las leyes lógicas descriptivas y que éstas a su vez se justifican apelando a las leyes lógicas descriptivas más básicas. Si estas afirmaciones reflejan correctamente el pensamiento de Frege, entonces Kusch (1995, 36) está en lo cierto cuando sostiene que la justificación de las leyes lógicas descriptivas básicas no puede ser de tipo lógico sino que parece necesario acudir a premisas psicológicas o biológicas.

5. LA PRIMACÍA DE HUSSERL Y EL FIN DEL DEBATE SOBRE EL PSICOLOGISMO

Pese a que el antipsicologismo de Frege es anterior al de Husserl, ya que sus primeras críticas explícitas datan de 1884, que su crítica al tratamiento de las entidades matemáticas de Husserl en *Philosophie der Arithmetik* fue decisiva en el antipsicologismo de éste, y que escribió su trabajo filosófico *Der Gedanke* en las postrimerías del debate sobre el psicologismo, en la filosofía académica alemana Frege permaneció casi ignorado. Por el contrario, Husserl pasó a ser considerado el filósofo que refutó definitivamente al psicologismo y su fenomenología se instaló raudamente en la comunidad filosófica alemana. Más aún, a pesar de que Husserl no realizó ninguna contribución específica a la lógica como ciencia, su fenomenología tuvo una gran influencia en el pensamiento lógico posterior, originando una concepción idealista de la lógica que se plasmó en los manuales de principios del siglo XX, entre los que cabe citar principalmente el de A. Pfänder, titulado *Logik*, de 1921 y otros textos menores. Éstos se extendieron a otras latitudes y, aunque en otros países ya se había impuesto la nueva lógica matemática, en los de habla hispana rigieron el enfoque y la

enseñanza de la lógica hasta la segunda mitad del siglo XX. En ellos, la lógica es ambiguamente presentada como la ciencia teorética que investiga los objetos lógicos en sí mismos, es decir, en tanto objetos ideales, independientes del contenido psicológico, y se la define o bien como la ciencia de las formas del pensamiento, es decir, una ciencia formal a la que no interesa el contenido (Grau, 1937); o bien como la ciencia de los pensamientos en cuanto tales, es decir, si se los considera prescindiendo de aquellos elementos que, aunque relacionados estrechamente con los pensamientos, no son los pensamientos mismos (Romero-Pucciarelli, 1945).

Las razones de la primacía de Husserl frente a Frege en relación con el psicologismo y el posterior auge la fenomenología son, según Kusch, de distinta índole. En su ya citado libro dedica sustanciosas páginas (203-210) a explicar tales razones, algunas de las cuales consideramos interesante sintetizar a continuación. En primer lugar, la lógica matemática era vista por los filósofos puros alemanes como un peligro tan grande como la psicología experimental. En efecto, muchos de ellos, particularmente Rickert y posteriormente Heidegger, veían en la logística un peligro tan grande como en el psicologismo. Este hecho explica el frío recibimiento que tuvo en la academia alemana el álgebra de la lógica de Boole en la última década del siglo XIX, rechazo éste que tuvo vigencia hasta muy entrado el siglo XX. Filósofos puros como Rickert, Natorp, Heidegger y el mismo Husserl, aunque con opiniones divergentes acerca de la valoración de la nueva lógica, coincidían en que la lógica matemática era una ilegítima incursión de la matemática en el dominio propio de la filosofía, de la misma manera que lo había sido la psicología experimental. Para ellos, la matemática y la lógica pertenecían a categorías ontológicas diferentes y por ello también debían ser consideradas ciencias independientes. En particular, el mismo Husserl se encarga de argumentar, en su comentario sobre la obra lógica de Schröder (*Vorlesungen über die Algebra der Logik*, 1890), a favor de una división tajante entre lógica y matemática, porque, mientras la lógica matemática no era una empresa filosófica, sí lo era la indagación acerca de la naturaleza de las entidades matemáticas. En segundo lugar, Frege era un pensador en soledad (la mayor parte de sus contribuciones y de su actividad académica la había realizado en la por entonces desprestigiada Universidad de Jena) y, respecto de la independencia de la filosofía frente a la matemática, no había escrito palabra alguna. A diferencia de Frege, Husserl era un conspicuo miembro de la comunidad académica de la época, la cual abiertamente optó por su posición idealista frente a la lógica y, en especial, por su propuesta de fundamentar la

filosofía en la lógica ideal concebida en su fenomenología. Por último, los estilos de argumentación de Frege y Husserl eran muy diferentes, ya que mientras Frege dedica sólo algunas páginas de sus obras a la crítica de la lógica psicológica y piensa que la mejor refutación es el desarrollo exitoso del programa logicista, Husserl dedica todo un libro al tema y polemiza con la mayoría de los defensores del psicologismo, atribuyéndose además a sí mismo la crítica a toda forma de psicologismo.

Por último, cabe destacar que el comienzo de la primer guerra mundial en 1914 hizo que las rivalidades dentro de la vida académica alemana cesaran y todos se agruparan bajo las consignas del pensamiento único que las circunstancias políticas imponían. Esto trajo a su vez consecuencias en la vida académica práctica, porque mientras la mayoría de los filósofos puros se centraron en el estudio de los aspectos ideológicos del «genio de la guerra» alemán, los psicólogos experimentales se dedicaron al entrenamiento y prueba de los soldados (Kusch, 1995, 211 ss.). Pero a su vez, ambos debieron enfrentarse o acomodarse al clima intelectual irracional de la época, totalmente hostil a la ciencia y a cualquier tipo de conocimiento racional y sistemático. En general, los psicólogos derivaron sus estudios a la psicología aplicada en las universidades técnicas, pero casi todos los que se quedaron en la vida académica optaron por rechazar la psicología experimental y abrazaron abiertamente una forma de psicología afín a la fenomenológica de Husserl. No quedaba, obviamente, espacio intelectual alguno para el programa de Frege.

Finalizada la primera guerra mundial, la polémica sobre el psicologismo no retornó al debate académico por dos razones fundamentales. Por un lado, porque la fenomenología de Husserl se consolidó en la filosofía académica alemana, y por el otro, porque la actitud antipsicologista de los filósofos puros hacia la psicología fue disminuyendo al punto de que puede afirmarse que en el año 1920 el debate sobre el psicologismo había desaparecido.

6. NUEVOS ENFOQUES

El primer enfoque realizado desde una perspectiva diferente de la tradición filosófica es el que sostiene Jean Piaget (1896-1980), fundador de la llamada epistemología genética. Piaget ha tratado las relaciones entre la lógica y la psicología en muchas de sus obras, pero hay una, escrita en forma conjunta con el matemático E. W. Beth, titulada *Epistemologie mathématique et psichologie. Essai sur les relations entre la*

logique formelle et la pensée reelle (Epistemología matemática y psicología. Relaciones entre la lógica formal y el pensamiento real) (1968), dedicada precisamente a esta temática. En ella se pretende determinar la total autonomía de la lógica en tanto ciencia demostrativa, respecto de la psicología, pero reivindicando la necesidad de mantener una relación entre ambas disciplinas a los efectos de constituir una epistemología científica. En el capítulo 7, denominado *Las lecciones de la historia de las relaciones entre lógica y psicología*, Piaget afirma:

> Pues, en la medida en que la lógica se ha orientado hacia el análisis de los fundamentos y de las condiciones de validez, no puede hacer otra cosa que desligarse de cualquier consideración de hechos; pero la psicología estudia el pensamiento en tanto que sistema de hechos, en su contexto causal, y ello aun cuando el sujeto se confine, en su conciencia, a consideraciones normativas. Por otra parte, en la medida en que la psicología se ha orientado hacia tal estudio de hechos, y por más que el psicólogo obedezca a sus métodos y normas y haya de plegarse a reglas lógicas, no podría, recíprocamente, sino desligarse de la lógica, ya que ninguna consideración de validez deductiva basta para zanjar problema alguno de hechos, que depende exclusivamente de la experiencia.

Sin embargo, pese a los innumerables esfuerzos de Piaget por mostrar la independencia entre la lógica y la psicología, ha sido tildado de logicista por los psicólogos y de psicologista por los lógicos. Para comprender esta doble crítica conviene recordar que uno de los puntos centrales de la epistemología genética lo constituye la teoría de los llamados *modelos lógicos*. En ella las formas de conocimiento que corresponden a cada etapa o período del desarrollo del conocimiento se describen por medio de una estructura (no matemática) pero formulada en términos lógico-algebraicos. Por ello Piaget llama *lógica operatoria* a la disciplina que tiene por finalidad describir tales estructuras junto a sus propiedades. Esta disciplina, aun cuando su herramienta principal sea el simbolismo de la lógica y del álgebra, tiene una finalidad psicológica, ya que los modelos que propone se pretende que describan los aspectos lógico-formales del pensamiento y que tengan prueba empírica. En otras palabras, la finalidad de los modelos lógicos o estructuras formales es describir el comportamiento cognoscitivo natural de los sujetos reales y por ello la lógica operatoria es una disciplina psicológica. Es precisamente la finalidad psicológica de la lógica operatoria la responsable de que su epistemología haya recibido el mote de psicologismo. Y, viceversa, que su teoría pretenda describir estados cognoscitivos apelando a estructuras algebraicas, es lo

que ha inducido a muchos psicólogos a interpretar a Piaget como sosteniendo que tales estructuras algebraicas están en la mente del sujeto epistémico y considerarlo por ello un auténtico logicista en relación con la psicología. Contrariamente a lo que sostienen estas interpretaciones, Piaget intenta mostrar que su epistemología presupone la independencia de ambas disciplinas, pero al mismo tiempo proclama, como ya lo hemos dicho, la necesidad de una coordinación entre ambas, apoyándose en las siguientes razones: por un lado, porque el conocimiento lógico matemático, en tanto normativo e interesado en las cuestiones de validez, se constituye en los sujetos reales como una invariante cognoscitiva, y por el otro lado, porque este mismo conocimiento ha tenido su génesis en las acciones de los sujetos reales. Esto es así porque, para la epistemología genética, del hecho de que tenga sentido la distinción entre ciencias formales (o normativas) por un lado y empíricas (o fácticas) por el otro, no se infiere que el origen del conocimiento específico de ambas haya de buscarse para las primeras en la razón y para las segundas en la experiencia, sino que ambas tienen su génesis en las acciones del sujeto respecto del mundo.

La segunda perspectiva que reseñaremos proviene también de un campo ajeno a la filosofía tradicional: la llamada Inteligencia Artificial (IA). En efecto, hacia fines de 1970 se instaló dentro del campo de la IA un intenso interés por la lógica, promovido por la necesidad de construir formalismos que pudieran representar en una máquina inteligente la capacidad humana de razonar, es decir, de representar el conocimiento de un sujeto real mediante un mecanismo simbólico que pudiera procesar distintas clases de símbolos mediante reglas formales (McCarthy y Hayes, 1969). Sin embargo, los especialistas en el tema inmediatamente se percataron de que el lenguaje y poder inferencial de la lógica clásica no se adecuaba para tal uso, ya que presentaba grandes discordancias con el tipo de razonamiento de los sujetos reales expresados en el lenguaje natural (Minsky, 1975 y Levesque, 1986). En particular, coincidieron en que el condicional material de la lógica clásica no era adecuado para representar las oraciones condicionales del lenguaje natural, porque mientras que en el primero el antecedente es siempre condición suficiente del consecuente, en los condicionales del lenguaje natural esto no se cumple. Por ejemplo, dada la oración *Si el autobús parte a las 17 horas de la tarde, entonces llegaré a destino a las 21 horas,* e interpretándola mediante el condicional material, no se podría dar el caso de que el autobús saliera a las 17 horas y no llegara a destino a las 21. Sin embargo, es natural suponer que, pese a salir a la hora señalada, el autobús no llegue a des-

tino a la hora prevista, porque, por ejemplo, haya sufrido un desperfecto en el motor. Luego, pese a que sea verdad la oración *El autobús partió a las 17 horas de la tarde*, no sea verdadera la oración *El autobús llegó a destino a las 21 horas.* En síntesis, el agregado de nueva información al antecedente de la oración condicional hace que el consecuente se torne falso y sea verdadera la oración: *Si el autobús parte a las 17 horas de la tarde pero sufre un desperfecto en el motor, entonces no llegará a destino a las 21 horas.* Los condicionales de este tipo han recibido, por las razones explicitadas, el nombre de condicionales *derrotables* o *revocables.* Pero además, es sabido que el condicional material es solidario de la noción de consecuencia lógica clásica, una de cuyas propiedades esenciales es la llamada *monotonía*, que es precisamente la que justifica en una inferencia válida agregar o reforzar la información de las premisas sin que ello influya en la verdad de la conclusión. Esto produjo que los lógicos dirigieran sus investigaciones hacia la caracterización de una nueva noción de consecuencia que reprodujera más fielmente las inferencias del lenguaje natural o del sentido común, la cual recibió el nombre de consecuencia no-monótona (Kraus, Lehmann y Magidor, 1990, y Makinson, 1994).

Según algunos autores (Pelletier y Elio, 1997), la creación de formalismos que intenten reproducir las formas de los razonamientos del sentido común conduce necesariamente a relacionar estos formalismos con las formas de razonar de los agentes reales y por ende a estudiar la normatividad propia de los mismos. Proponen para ello una especie de psicologismo que postula considerar como válidos aquellos razonamientos que el agente común considera como tales, es decir, las inferencias no monótonas aceptadas como correctas por los agentes. Desde una perspectiva diferente y en general provenientes de la psicología cognitiva, varios autores suponen que en la práctica natural de los agentes hay una normatividad o competencia propia, llamada comúnmente lógica natural o lógica del sentido común y en la que actualmente se investiga a fin de construir modelos lógicos que la representen (Braine, 1990, y Moshman, 1990). Lejos de replantear el problema en términos del psicologismo clásico, no queda duda alguna de que estos enfoques nos retrotraen a la problemática de las relaciones entre lógica y psicología y, en particular, a la distinción de Frege entre leyes lógicas descriptivas y leyes lógicas prescriptivas y a la pregunta acerca de la génesis de las inferencias lógicas, respecto de la cual la epistemología genética ensayó una primera respuesta.

	Tradición psicologista		Tradición logicista
1520	Introducción del término *Psicología* por Goclenius.		
1635	*Logica Hamburgensis*, J. Jungius	Enfoque psicológico de	
1662	*Logique de Port Royal. La logique ou l'art de penser*, de A. Arnauld y P. Nicole.	la lógica	
1664			*Ars Combintoria*, de Leibniz
1807	*Neue Kritik der Vernunft*, de J. Fries		
1832	*Lehrbuch der Logik als Kunstlehre des Denkens*, de F. Beneke	Naturalismo o psicologismo filosófico	
1837			*Wissenschaftslehre*, de B. Bolzano
1842	*System der Logik als Kunstlehre des Denkens*, de F. Beneke		
1843	*System of Logic*, de J. S. Mill		
1847			– *Mathematical Analysis of Logic*, de G. Boole – *Formal Logic*, de A. de Morgan
1854			*An Investigation of the Laws of Thought*, de G. Boole
1870		Surgimiento de la nueva psicología	
1873	*Logik*, de C. de Sigwart		
1878			*Begriffsschrift*, de G. Frege
1883	*Logik*, de W. Wundt		
1884			*Die Grundlagen der Arithmetik*, de G. Frege
1886	Introducción del término *Psicologismus* por E. Erdmann	Debate	
1893	*Grundzüge der Logik*, de T. Lipps	sobre	
1893-1903		el	*Grundgesetze der Arithmetik*, de G. Frege
1894		psicologismo	Reseña de Husserl, *Philosophie der Arithmetik*, en *Zeitschrift für Philosophie und Philosophische Kritik*, de G. Frege
1900-1901	*Logische Untersuchungen*, de E. Husserl		
1914-1918		Primera guerra mundial	
1918			«Der Gedanke», de G. Frege
1920		Extinción del debate sobre el psicologismo	

BIBLIOGRAFÍA

Arnauld, A. y Nicole, P. (1851), *La Logique de Port Royal*, edición sin pie de imprenta que forma parte de la Colección Mielli de la Biblioteca Central de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires.

Ashworth, E. J. (1998), *Logic, Renaissance*, en *Routledge Encyclopedia of Philosophy* (*REPh*), versión 1.0, Routledge, London/New York.

Beth, E. W. y Piaget, J. (1968), *La relación entre la lógica formal y la psicología,* Ciencia Nueva, Madrid.

Bochenski, I. M. (1976), *Historia de la Lógica Moderna*, Gredos, Madrid.

Braine, M. D. S. (1990), «The natural logic approach to logic», en W. F. Overton (ed.), *Reasoning, necessity and logic,* Lawrence Erlbaum Associates, New Jersey/London.

Brewka, G. (1991), *Non-Monotonic Reasoning: Logical Foundations of Common sense,* Cambridge University Press, Cambridge.

Capozzi, M., «Logic in the 17th and 18th centuries»: *REPh.*

Carnotta, R. (1995), «Lógica e Inteligencia Artificial», en *Enciclopedia Ibero-Americana de Filosofía*, vol. 7, *Lógica*, Trotta-CSIC, Madrid.

Coffa, A. (1991), *The Semantic tradition from Kant to Carnap. To the Vienna Station*, Cambridge University Press, Cambridge.

Dipert, R. R. (1998), «Logic in the 19th century», en *REPh.*

Elio, R. y Pelletier, J. (1994), «On relevance in nonmonotonic reasoning: Some empirical studies», en R. Greiner y D. Subramanian (eds.), *Relevance*, American Association For Artificial Intelligence, Symposium Series, November 4-6, New Orleans, 64-67.

Føllesdal, D. (1969), «Husserl's Notion of the Noema», en *Journal of Philosophy*, vol. 66, 680-687.

Føllesdal, D. (1995), *Gödel and Husserl*, en J. Hintikka (ed.), *From Dedekind to Gödel: Essays on the Development of the Foundations of Mathematics*, Kluwer, Dordrecht.

Føllesdal, D. (1998), *Edmund Husserl*, en *REPh.*

Frege, G. (1879), *Begriffsschrift, eine der arithmetisch nachgebildete Formelsprache des reinen Denkens*; traducción inglesa: «Begriffsschrift, a Formula Language Modelled upon that of Arithmetic, for Pure Thought», en J. van Heijenoort (ed.), *From Frege to Gödel: A source book in Mathematical Logic*, Harvard University Press, Cambridge, MA, 1967.

Frege, G. (1884), *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*; trad. inglesa de J. L. Austin: *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Blackwell, Oxford, 1980.

Frege, G. (1893-1903), *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*; traducción parcial al inglés de M. Furth: *Basic Laws of Arithmetic: An Exposition of the System*, University of California Press, Berkeley, 1964.

Frege, G. (1918), «Der Gedanke: eine logische Untersuchung»; traducción inglesa en B. McGuinnes (ed.): *Collected Papers on Mathematics, Logic and Philosophy*, Blackwell, Oxford.

Gabbay, D. M., Hogger, C. J. y Robinson, J. A. (eds.) (1994), *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3: *Nonmonotonic Reasoning and Uncertain Reasoning*, Oxford University Press, Oxford.

Grau, K. J. (1937), *Lógica*, Labor, Barcelona/Madrid.

Haack, S. (1979), *Philosophy of Logics*, Cambridge University Press, London/New York.

Horty, J. (1998), «Common Sense, Theories»: *REPh*.

Husserl, E. (1967), *Investigaciones lógicas*, Revista de Occidente, Madrid.

Kneale, W. y M. (1972), *El desarrollo de la lógica*, Tecnos, Madrid.

Kraus, S., Lehmann, D. y Magidor, M. (1990): «Nonmonotonic Reasoning, Preferential Models and Cumulative Logics»: *Artificial Intelligence,* 44, 167-201.

Kusch, M. (1995), *Psychologism. A case study in the sociology of philosophical knowledge,* Routledge, London/New York.

Lalande, A. (1916), *Les rapports de la logique et de la psychologie, Scientia*, vol. XX, Bologna/London/Paris.

Levesque, H. (1986), «Knowledge Representation and Reasoning»: *Annual Reviews of Computer Science*, 1, 255-287.

MacNamara, J. (1986), *A border Dispute: The Place of Logic in Pshychology*, MIT Press, Cambridge, MA.

Makinson, D. (1994), «General Patterns in Non-Monotonic Reasoning», en D. M. Gabbay, C. J. Hogger y J. A. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Non-Monotonic Reasoning*, Oxford University Press, Oxford.

Mates, B. (1974), *Lógica matemática elemental*, Tecnos, Madrid.

McCarthy, J. y Hayes, P. (1969), «Some Philosophical Problems from the Standpoint of Artificial Intelligence», en Meltzer y Michie (eds.), *Machine Intelligence*, 4, 463-502, Edinburgh University Press, Edinburg.

Mill, J. S. (1967), *System of Logic*, Longmans, United Kingdom.

Minsky, M. (1975), «A Framework for Representing Knowledge», en *The Psychology of Computer Vision*, McGraw-Hill, New York.

Moshman, D. (1990), «Development of Metological Understanding», en Overton, W. F. (ed.): *Reasoning, necessity and logic*, Lawrence Erlbaum Associates, New Jersey/London.

Olafson, F. (1975), *Husserl Theory of Intentionality in Contemporary Perspective*, en *Noûs*, 9, 73-83.

Overton, W. F.: «Competence and Procedures: Constraints on the Development of Logical Reasoning»: *RNL*, 2-31.

Pelletier, F. J. y Elio, R. (1997), «What should Default Reasoning be by Default?»: *Computational Intelligence*, 13, 165-187.

Romero F. y Pucciarelli, E. (1945), *Lógica*, Espasa-Calpe, Buenos Aires.

IDENTIDAD Y REFERENCIA

Rodolfo Gaeta

1. DIMENSIONES ONTOLÓGICA, LÓGICA Y EPISTEMOLÓGICA DEL CONCEPTO DE IDENTIDAD

A propósito de la identidad surgen dos grupos de problemas distintos. Por una parte, la identidad puede entenderse como el concepto que alude a la permanencia de una cosa a través de sus cambios. Una segunda clase de problemas se vincula con la posibilidad de que haya dos o más objetos que compartan estrictamente todas sus propiedades, es decir, la posibilidad de que cosas idénticas sean numéricamente diferentes; el análisis de esta cuestión es independiente del tema de la permanencia porque, aun suponiendo que nos ocupáramos solamente de entidades inmutables, cabe preguntarse cómo caracterizar su identidad en oposición a la diversidad numérica.

La identidad puede concebirse como una peculiar relación que cada entidad tiene consigo misma; en ese caso, aparece como un rasgo primordial de la realidad y queda incorporada al plano ontológico o metafísico. Es frecuente también presentarla como un principio de la lógica. Por último, es posible sostener que las leyes de la identidad corresponden a la esfera del conocimiento.

Aunque la distinción de estas dimensiones —ontológica, lógica y epistemológica— reviste gran relevancia, la circunstancia de que las leyes de la identidad puedan considerarse válidas en cualquiera de ellas hace que en muchas ocasiones se las formule sin especificar su dominio. En otros casos, esa indicación aparece de manera indirecta, como cuando se supone que las leyes de la lógica reflejan lo que primariamente rige en la esfera ontológica, o cuando se establece que la lógica fija las pautas del conocimiento. Con frecuencia la formulación del principio de identidad funde, o confunde, estos aspectos.

La creencia ampliamente difundida de que, de todos modos, la identidad tiene reservado un lugar en la lógica contrasta con la opinión que mantienen William y Martha Kneale. Estos autores conciben la lógica formal como una teoría sobre la inferencia o el desenvolvimiento de proposiciones que se formula por medio de reglas de inferencia, al estilo de Gentzen, en lugar de axiomas. Y como creen que no pueden incluirse reglas del tipo adecuado que sean capaces de reproducir el papel de los axiomas de la identidad, concluyen que esta noción queda fuera de la lógica (Kneale, 1992).

Aristóteles, que formuló las leyes básicas de la identidad, advirtió que la utilidad de los enunciados que expresan identidades deriva del hecho de que el lenguaje permite referirse a una misma cosa con nombres diferentes. Así, sostiene que llamamos idéntico según el número a lo que tiene más de un nombre, como vestido y manto, aunque en realidad se trata de una sola cosa. La formulación aristotélica de los principios lógicos muestra también la superposición de puntos de vista diferentes.

La distinción entre el aspecto gnoseológico y el ontológico aparece con toda claridad en la filosofía medieval. Los pensadores medievales consideraban que la lógica, cuyo papel como instrumento del conocimiento está fuera de discusión, tiene por objeto las *segundas intenciones*, entes de razón que no se confunden con los entes naturales. Santo Tomás lo explicita en estos términos:

> La relación designada por la palabra «mismo» es puramente de razón, si tal palabra se toma sin más, porque una tal relación no puede consistir más que en una ordenación que halla el entendimiento de una cosa a sí misma, considerada dos veces (Tomás de Aquino, *Sum. Theol.*, I, 28, 1, ad. 2).

De acuerdo con la opinión de Stroll (Stroll, 1972) casi todos los filósofos modernos, desde Descartes hasta Kant, reconocieron en el principio de identidad una ley básica del pensamiento, que hoy expresaríamos con la fórmula «(x) (x = x)». De todos modos, y a pesar del rigor con que el problema había sido planteado en la época medieval, muchas veces es difícil encontrar manifestaciones explícitas que permitan determinar si el sentido primario que se atribuye a la identidad es lógico u ontológico. Una muestra de las dificultades que entraña esta tarea se encuentra en la filosofía de Kant. Roberto Torretti indica que a veces Kant se refiere a los principios lógicos como si fueran uno solo, porque sostiene que se implican mutuamente. Pero lo más importante es que tales principios no se acomodan a la distinción entre juicios analíticos y juicios sintéticos. No podrían ser sinté-

ticos, obviamente, a menos que se tratara de sintéticos *a priori*. Pero la explicación kantiana sobre la posibilidad de los juicios sintéticos *a priori* no puede aplicarse a esta clase de principios. Tampoco parece tratarse de juicios analíticos, en vista de que Kant atribuye a la no contradicción y la identidad el papel de un principio o norma de todos los juicios analíticos pero le niega el carácter de axioma y lo describe como una fórmula o modelo universal de los juicios analíticos (Torretti, 1980, 236 n.).

La confusión del aspecto lógico con el ontológico se manifiesta especialmente en el idealismo postkantiano, en cuyas doctrinas la lógica aparece absorbida por la metafísica al punto tal que inclina a pensar que, en cualquier sentido formal, la lógica ha desaparecido.

La reivindicación del carácter estrictamente lógico de esta clase de principios ha corrido por cuenta de los neopositivistas y es consecuencia tanto de su revalorización del papel de la lógica como de su manera de concebirla. Carnap declara al respecto:

> Puesto que todas las proposiciones de la lógica son tautológicas y vacías de contenido, no podemos inferir de ellas lo que sea necesario o sea imposible. Así pues, el intento de basar la metafísica en la lógica, que es característica principalmente de un sistema como el de Hegel, resulta injustificado (Carnap, 1965, 158).

Asistimos así a una curiosa sustitución. Los empiristas lógicos reaccionaron en contra de todo intento de vincular la ontología con la lógica, pero terminaron reduciendo esta última a una suerte de capítulo de la gramática.

2. LA CONCEPCIÓN DE FREGE ACERCA DE LOS ENUNCIADOS DE IDENTIDAD

En su artículo «Sobre el sentido y la denotación», Frege comienza por preguntarse si la identidad es una relación, y en tal caso, si se trata de una relación entre objetos o entre nombres de objetos. Sostener que la identidad es una relación entre nombres permitiría superar una dificultad llamada a veces, precisamente, «paradoja de Frege». La dificultad salta a la vista cuando se compara un enunciado de la forma «$a = a$» con otro de la forma «$a = b$». Si la identidad fuese una relación entre objetos, bajo el supuesto de que efectivamente a y b son idénticos, ambos enunciados debieran decir exactamente lo mismo. Sin embargo, mientras que «$a = a$» es, en términos de Kant, analítico y *a priori*, las oraciones de la forma «$a = b$» constituyen muchas ve-

ces una ampliación del conocimiento que no puede obtenerse *a priori* sino a través de la investigación empírica, como la que llevó a la conclusión de que el lucero matutino es idéntico al lucero vespertino. Para explicar la diferencia en cuanto al valor cognoscitivo de ambos enunciados, Frege había llegado a sugerir que «$a = b$» pretende decir que «a» y «b» son dos nombres que denotan un mismo objeto, de modo que el enunciado expresa una relación entre signos y resultaría verdadero en la medida en que «a» y «b» fueran usados realmente para designar el mismo objeto. Pero como los nombres son arbitrarios, en tanto dependen de convenciones lingüísticas, sostener que a es idéntico a b no representaría nunca, aunque fuese verdad, la expresión de un conocimiento genuino. Semejante consecuencia choca, evidentemente, con el hecho de que muchos enunciados de identidad tienen un real valor cognoscitivo.

Pero la idea de que la identidad sea una relación entre nombres ha sido rechazada por Saul Kripke conforme al siguiente argumento:

> Si uno dice que para todo x y para todo y, si $x = y$ entonces $y = x$, o algo por el estilo, en dicho enunciado no figura ningún nombre y nada se dice acerca de nombres. Este enunciado sería verdadero aunque la raza humana nunca hubiese existido, o aun si existiera y nunca se hubiese producido el fenómeno de los nombres (Kripke, 1981, 107).

Sin duda, Frege no negaría que la identidad de las cosas valdría aun cuando no hubiese existido la humanidad, pero él estaba interesado en estudiar las propiedades de los enunciados de identidad, a saber, su valor cognoscitivo. Y concluyó que sólo cuando los nombres que flanquean el signo de identidad poseen sentidos diferentes el enunciado no resulta trivial. Esto no quiere decir que tales enunciados *se refieran* a los sentidos, sino que las diferentes designaciones de un mismo objeto indican, a la vez, el modo de presentación, y es por ello que la oración expresa un conocimiento genuino.

Pero hay consecuencias de la paradoja de la identidad que no se refieren al valor cognoscitivo sino al valor de verdad. Si en el enunciado «La estrella matutina es la estrella matutina» aplicamos el principio de sustitutividad de los idénticos —sobre la base de que ambos nombres designan, de hecho, el mismo objeto— y lo transformamos en «La estrella matutina es la estrella vespertina», la verdad se conserva; pero si hacemos la misma sustitución en el enunciado «Juan cree que la estrella matutina es la estrella matutina», el valor veritativo podría alterarse. Lo mismo sucedería con el enunciado «Necesariamente la estrella matutina es la estrella matutina», si una de las dos

apariciones de «estrella matutina» es reemplazada por «estrella vespertina», puesto que la identidad se tornaría contingente.

3. NECESIDAD *DE DICTO* Y NECESIDAD *DE RE*

Resultaría muy extraño negar la afirmación de que toda cosa es necesariamente idéntica a sí misma. No obstante, sólo una investigación empírica permite establecer la verdad de enunciados tales como «La estrella matutina es la estrella vespertina», de manera que no parecen ser sino contingentes.

Esta problemática nos obliga a considerar el concepto de necesidad. En primer lugar, cabe preguntarse si la necesidad reside en la propia realidad o en los juicios que formulamos acerca de ella. Los filósofos medievales distinguieron, precisamente, los dos casos. Una oración —señalaron— en la cual figura un término modal, como «necesariamente» o «posiblemente», puede interpretarse, en principio, de dos maneras: 1) en el sentido compuesto (*sensus compositus*), de modo que el término modal se toma como un predicado de segundo orden aplicado a una proposición asertórica (*el dictum*); 2) en el sentido dividido (*sensus divisus*), en cuyo caso se denomina también modalidad *de re*, donde la partícula modal ya no se predica de un *dictum*, sino que modifica la cópula, o, si se quiere, el predicado de la proposición asertórica.

De acuerdo con la interpretación *de dicto*, la oración «Necesariamente todos los hombres son racionales» expresa en realidad una proposición metalingüística, y según las convenciones ahora acostumbradas podría expresarse así: «'Todos los hombres son racionales' es una verdad necesaria». Si se entiende, en cambio, que la proposición original tiene sentido dividido, o *de re*, cabría formularla con más claridad de esta otra manera: «Todos los hombres son necesariamente animales». Pero debe observarse que este enunciado conserva aún cierta ambigüedad que no hace totalmente descartable una interpretación que le asigne sentido compuesto. En el moderno simbolismo de la lógica modal la ambigüedad desaparece: si el operador modal figura antepuesto al cuantificador universal, expresa una modalidad *de dicto*; si está dentro del alcance del cuantificador, se trata de una modalidad *de re*.

El valor veritativo de un enunciado modal puede variar conforme a cuál de las interpretaciones se acepte. Consideremos, siguiendo a Quine, el enunciado «Necesariamente un competidor ganará» y supongamos que se formula con respecto a cierto juego de competido-

res individuales que no admite empates. El enunciado podría interpretarse tanto en sentido compuesto como en sentido dividido. Si se interpreta en el sentido compuesto, lo que se afirma es simplemente que por lo menos un jugador y no más de uno ganará, pero no hay un determinado jugador del que pueda decirse que ganará. Quien atribuya al enunciado sentido dividido interpretará, en cambio, que hay cierto jugador, por más que el enunciado no lo identifique, que necesariamente resultará vencedor. Se advierte que ambas interpretaciones no son equivalentes si se observa que en caso de que el enunciado resulte verdadero en la interpretación *de re* también lo sería de la otra forma, pero la recíproca no es válida, porque el que haya uno y no más de uno de los competidores que resultará ganador no implica que quien efectivamente resulte vencedor debiera ganar necesariamente.

Cabe preguntarse, por otra parte, con qué fundamentos puede predicarse la necesidad. Quine sostiene en «Three Grades of Modal Involvement» que la necesidad en cuanto predicado semántico atribuido a nombres de enunciados ni siquiera pertenece a la lógica modal, porque en la medida en que se construye dentro de una extensionalidad estricta forma parte de la teoría de la prueba, aunque tienda a confundírsela con la analiticidad. Esta tendencia complica la situación, porque el predicado «es analítico», en su sentido usual, no es extensional: no basta conocer el valor de verdad de un enunciado para determinar si es o no analítico. Las observaciones de Quine sugieren que la expresión «necesariamente» *no debe* interpretarse como una manera de decir que el enunciado al que se aplica es analítico, lo cual es coherente con sus proverbiales reservas respecto de la noción de analiticidad; su contrastante disposición a admitir el uso de términos modales referidos a contextos extensionales indica que no está pensando en ningún otro tipo de necesidad fuera de la rigurosa necesidad lógica.

Lo que Quine califica como «segundo grado» de compromiso modal es el caso en que las partículas modales funcionan como operadores lógicos que se aplican a enunciados —de igual forma que las conectivas corrientes— y no a nombres de enunciados. Quine está dispuesto a admitir también este uso de la noción de necesidad, siempre que se lo interprete como una versión estilística de la modalidad de primer grado; pero no lo acepta si el operador correspondiente se emplea para formar enunciados no extensionales. Considera lícito, por ejemplo, formular el enunciado «Necesariamente (9 es mayor que 5)» sólo si se lo interpreta como una manera de expresar «Necesariamente '9 es mayor que 5'». Si no se adopta este recaudo, quedaría abierta la posibilidad de sustituir «9» por «el número de los planetas»

y el enunciado resultante, «necesariamente (el número de los planetas es mayor que 5)» ya no sería verdadero.

Finalmente, el tercer grado de compromiso modal extiende la aplicabilidad del operador modal de manera que pueda anteponerse también a oraciones abiertas (*open sentences*). Esta posibilidad es la que corresponde a la lógica modal cuantificada, la versión modernizada de la teoría de las modalidades *de re*. Si esto es así, ¿cuál es el lugar que corresponde en este esquema a las modalidades *de dicto*? ¿Coincide con la necesidad analítica? No podemos decir simplemente que la modalidad *de dicto* corresponde al primer grado del que habla Quine, pues, conforme a su criterio, este uso de las partículas modales no pertenece ni siquiera a la modalidad no cuantificada. Descartado el tercer grado, ya que por hipótesis aceptamos que representa las modalidades *de re*, sólo resta la alternativa de identificar la auténtica necesidad *de dicto* con lo que Quine denomina segundo grado, si es que no queremos negar absolutamente la existencia de modalidades *de dicto*. Con todo, nos apartaríamos de la idea de los lógicos medievales, quienes entienden que la modalidad *de dicto* se expresa a través de enunciados de segundo orden o metalingüísticos. Estas observaciones sugieren, pues, que las modalidades de *dicto*, en cuanto pertenezcan a la auténtica lógica modal, ocupan un segundo grado (siempre que no se trate de una paráfrasis de las de primer grado).

En lo que concierne a la relación que existe entre las modalidades y la analiticidad, Quine escribe:

> La idea general de modalidades estrictas se basa en la muy putativa idea de analiticidad del modo siguiente: un enunciado de la forma «Necesariamente...» es verdadero si y sólo si el enunciado componente regido por «necesariamente» es analítico; y un enunciado de la forma «posiblemente...» es falso si y sólo si la negación del enunciado componente regido por «posiblemente» es analítico (Quine, 1963, 206-207).

En esta ocasión Quine está dispuesto a admitir, en beneficio de la argumentación, que las modalidades tienen sentido, aunque ello implique dejar de lado momentáneamente sus reservas con respecto a la analiticidad, pero sigue oponiendo objeciones a la cuantificación lisa y llana de la lógica modal, que juzga carente de sentido (*ibid.*, 125).

Algunos lógicos rechazan la legitimidad de la distinción entre modalidad *de dicto* y *de re* sin compartir la desconfianza de Quine acerca de las modalidades en general. Von Wright, por ejemplo, piensa que se puede prescindir de la cuantificación de contextos modales porque estima que todo lo que se pretende expresar mediante esa

cuantificación puede formularse igualmente por medio de enunciados en los cuales el operador modal no figura en el alcance de un cuantificador. Esta sugerencia se funda en la distinción entre las propiedades *formales* y las propiedades *materiales* de los objetos, conforme a lo que von Wright llama *Principio de Predicación*. Los atributos formales, entre los cuales se hallan las propiedades lógicas y aritméticas, pueden predicarse indistintamente por medio de la lógica modal cuantificada o no cuantificada. Las propiedades materiales, en cambio, serían todas contingentes y quedan fuera del ámbito de la necesidad (von Wrigth, 1951, 26-28).

Pero W. y M. Kneale señalan que von Wright no parece haber tomado en cuenta que el hecho de que un atributo pertenezca necesariamente al sujeto no depende exclusivamente de la naturaleza de dicho atributo, porque puede depender también del sentido del término sujeto.

> Si se supone que la cualificación *de re* excluye la dependencia de la modalidad respecto del sentido en cuanto diferente de la referencia del término-sujeto habrá que concluir paradójicamente que los únicos enunciados modales *de re* verdaderos son aquellos cuyo término-sujeto puede ser reemplazado *salva veritate* por cualquier otra cosa de la misma categoría, esto es, enunciados tales como «La reina Ana está necesariamente muerta o no muerta», en que el término-sujeto carece de importancia en orden a determinar la verdad del enunciado en su conjunto y ésta depende simplemente del carácter tautológico de la expresión «muerta o no muerta» (Kneale, 1992, 225).

Estos autores opinan que, aparentemente, von Wright se ha dejado confundir por la «dudosa distinción» entre modalidades *de dicto* y *de re*. Pero concluyen, de todos modos, fundándose en argumentos propios, que la cuantificación modal es prescindible.

Por otra parte, la posibilidad de que existan modalidades *de re* se ha planteado tradicionalmente en el terreno metafísico. Pero hay, en principio, otras alternativas. Quine está dispuesto a aceptar alguna forma de necesidad natural (Quine, 1965, 76), y aunque no desarrolla la idea, alude a la probable conveniencia, y aun a la necesidad, de cuantificarla en el discurso científico.

4. LA TEORÍA DE RUSSELL ACERCA DE LOS NOMBRES PROPIOS Y LAS DESCRIPCIONES DEFINIDAS

Todo intento de explicar las características de los enunciados de identidad se funda en una teoría semántica. Ya hemos aludido a la solu-

ción de Frege. Según su doctrina, los nombres propios poseen a la vez denotación y sentido, y lo mismo ocurre con las descripciones definidas, del tipo «la estrella matutina». En un lenguaje perfecto cada nombre propio tendría un único sentido y toda descripción debe tener asegurada una denotación; en caso de carecer de ella corresponde fijarla convencionalmente, como se hace en matemáticas cuando se fija un valor arbitrario para ciertas funciones. Pero en los lenguajes corrientes un mismo nombre posee sentidos diferentes para distintas personas, así como muchas veces las descripciones no tienen denotación.

A partir de estas ideas, Frege brinda una solución a las paradojas de la identidad: las expresiones «$a = a$» y «$a = b$» pueden ser ambas verdaderas, pero la segunda posee un valor cognoscitivo del que carece la primera. Ello se explica porque, fuera del caso en que ya se sabía que «b» es un sinónimo de «a», «b» posee un sentido diferente del que corresponde a «a», y por lo tanto si la identidad «$a = b$» es verdadera, expresa un conocimiento genuino que requiere, en el caso de que a y b sean objetos materiales, una investigación empírica. En un ejemplo de este tipo «$a = b$» expresa un conocimiento que no puede ser considerado analítico, y la propia identidad formulada sería contingente. Por otra parte, si alguno de los nombres que figuran en el enunciado expresado en lenguaje directo no posee denotación, la oración completa carece también de denotación, es decir, no tiene valor de verdad. Naturalmente, lo mismo ocurre si la expresión que carece de denotación es una descripción.

Aunque en un primer momento Russell aceptó estas ideas, más tarde elaboró una teoría semántica muy ligada con sus propias tesis epistemológicas. Russell distingue el *conocimiento directo* (*by acquaintance*), en cuyo caso el objeto conocido está inmediatamente presente al sujeto y queda nombrado por un demostrativo como «éste», y el conocimiento *por descripción*, que constituye la mayor parte de nuestro saber, y que se refiere a objetos que no están directamente presentes. En el primer caso los demostrativos son auténticos nombres propios, nombres en sentido lógico. En el segundo se usa una descripción definida de la forma «El tal y tal...», la cual, en principio, alude a un objeto caracterizado por una propiedad que lo singulariza. En cuanto a los nombres propios ordinarios, abrevian o encubren una descripción del objeto denotado.

Las descripciones definidas tienen una estructura lógica que incluye la afirmación de la existencia y unicidad de cierto objeto. De acuerdo con la primera tesis «El actual rey de Francia es calvo», debe simbolizarse de la siguiente manera:

$$(\exists x)\ [Rx\ .\ Cx\ .\ (y)\ \ (Ry \supset y = x)]$$

De este modo, la oración en cuestión resulta ser falsa, lo mismo que su negación, si entendemos esta última de la manera adecuada, esto es:

$$(\exists x)\ [Rx\ .\ \sim Cx\ .\ (y)\ \ (Ry \supset y = x)]$$

La solución propuesta por Russell permite superar varias dificultades. En cuanto a los enunciados de identidad de la forma «A = B», donde «A» y «B» son nombres propios, se reducen a casos de identidades expresadas por medio de las dos descripciones abreviadas por «A» y «B» respectivamente y así se explica su valor informativo.

5. LA CONCEPCIÓN DE SEARLE SOBRE LOS NOMBRES PROPIOS: UNA SOLUCIÓN INTERMEDIA

Tanto Frege como Russell, pues, suscriben una concepción descriptivista de los nombres propios del lenguaje ordinario, pero no han logrado convencer a todos los filósofos. En su artículo «Proper names and descriptions», J. Searle examina los argumentos esgrimidos en contra de esa tesis y presenta su propia solución. Admite que los nombres propios no se usan para describir objetos, pero no niega que existe una conexión lógica entre el nombre y las características del objeto denotado. Searle estima que Mill estuvo en lo cierto al señalar que los nombres propios no implican ninguna descripción particular, pero cree que Frege también formuló una observación correcta cuando indicó que todo término singular debe expresar algún modo de presentación del objeto correspondiente; su error fue el de identificar la descripción que sirve para individualizar el objeto con una definición. La situación real —según Searle— es que hay una multiplicidad de descripciones *posibles* de cada objeto; aunque no es ninguna de ellas individualmente sino la disyunción que forman la que está analíticamente vinculada con el nombre. Si igualáramos los nombres a las descripciones, deberíamos elegir una descripción como aquella que el nombre representa. Pero entonces correríamos el riesgo de descubrir finalmente que dicha descripción no es satisfecha por el objeto, y en ese caso el denotado del nombre, si existe, es otro individuo. Para evitar esta consecuencia, que se aparta del uso efectivo de los nombres propios, hay que pensar que se dispone de otras descripciones alternativas que garantizan la conservación del mismo denotado. Así, pese

a que cualquier descripción sólo contingentemente será satisfecha por el objeto, su disyunción hace posible que el nombre se siga utilizando para hacer referencia al objeto del cual se intenta hablar.

6. LA NUEVA TEORÍA DE LA REFERENCIA

En la década de los años setenta se dio a conocer una nueva teoría que abandona la concepción descriptivista de los nombres y rescata las tesis de Mill. Se la ha denominado «la nueva teoría de la referencia» y habitualmente se considera que fue elaborada independientemente por Hilary Putnam y Saul Kripke. En los últimos años, sin embargo, se ha planteado una controversia desatada por un trabajo leído por Quentin Smith en una reunión de la American Philosophical Association en el cual sostuvo que las principales ideas de la nueva teoría habían sido formuladas por Ruth Barkan Marcus en el período transcurrido entre 1946 y 1961. El ensayo de Smith generó una réplica de Scott Soames, quien defendió la originalidad de las ideas de Kripke y a partir de allí se sucedieron otras discusiones al respecto[1]. De todos modos, en lo que sigue nos concentraremos en las ideas de Kripke, cuyas consecuencias resultan particularmente importantes en relación con el problema de la identidad.

Kripke sostiene que los nombres propios integran una clase de expresiones, a las que denomina «designadores rígidos», que se caracterizan por el hecho de que designan el mismo individuo en todos los mundos posibles en los cuales tal individuo existe, es decir, en cualquier condición contrafáctica concebible que incluya su presencia. Así, «Nixon» y «Venus» son designadores rígidos porque nombran, respectivamente a la misma persona y el mismo planeta en todos los mundos posibles en los que esté incluida su existencia, por más diferencias que los separen del mundo real.

La contrapartida de los designadores rígidos está constituida por los designadores accidentales: aquellas expresiones que sirven para designar pero no tienen por qué referirse al mismo individuo en todos los mundos posibles. Los ejemplos clásicos de estos designadores accidentales son la mayoría de las llamadas «descripciones definidas». Así, «El presidente de los Estados Unidos en 1970» no es, por cierto, un designador rígido, puesto que en otros mundos posibles, en otras condiciones contrafácticas, podría haber correspondido a Kennedy o a cualquier otra persona en lugar de referirse a Nixon.

1. Los artículos de Smith y Soames, conjuntamente con otros trabajos relevantes, han sido recopilados en Humphreys y Fetzer (comps.) (1998).

De acuerdo con los supuestos de Kripke, hay algunos individuos que existen en todos los mundos posibles. Tal es el caso de las entidades matemáticas. Atento a esta circunstancia, Kripke denomina «designadores débilmente rígidos» a los que designan un mismo individuo en todos los mundos posibles en los que ese objeto existe, cuando se trata de un individuo que no existe en todos los mundos posibles. «Nixon» es un designador de ese tipo, porque es concebible que podría no haber existido el individuo que existió en el mundo real y respondía a ese nombre. Un designador *fuertemente rígido*, en cambio, es el que designa un objeto cuya existencia es necesaria. Así, «9» es un designador fuertemente rígido, porque designa el mismo número en todos los mundos posibles.

Kripke reconoce que frecuentemente se apela a alguna descripción para introducir un nombre, especialmente cuando no es posible presentar el objeto correspondiente en forma ostensiva. Pero inmediatamente señala que la descripción sólo sirve para fijar la referencia; es decir, cumple un papel en una suerte de bautismo, pero una vez que el nombre es introducido pierde todo aspecto descriptivo. A juicio de Kripke, los hablantes pueden constatar la rigidez de los nombres a través de la intuición directa. Esta misma intuición básica es la que permite saber que nos referimos al mismo individuo cuando lo imaginamos en condiciones contrafácticas.

Pero no solamente los nombres propios ordinarios son designadores rígidos. Comparten este carácter los pronombres demostrativos y aun las variables libres. Lo más llamativo, quizás, es que Kripke extiende el alcance de la designación rígida a ciertos tipos de nombres generales. Es el caso de los que se refieren a clases naturales, los términos de masa («agua», por ejemplo) y los que designan fenómenos naturales, tales como «calor».

Kripke introduce una distinción ulterior. Denomina rigidez *de jure* a la que resulta de estipular que un nombre determinado corresponde a cierto objeto, ya sea en el mundo real o en cualquier situación contrafáctica, y la contrapone a la rigidez *de facto*, que aparece cuando una descripción definida contiene un predicado que es satisfecho en todo mundo posible por un mismo y único individuo (por ejemplo «el menor número primo», que designa siempre el número dos). De acuerdo con esta aclaración, los nombres propios ordinarios deben considerarse designadores rígidos *de jure*, mientras que algunas descripciones se comportan como designadores rígidos *de facto*.

7. CONSECUENCIAS DE LA SEMÁNTICA DE KRIPKE: UNA REFORMULACIÓN RADICAL DE LAS PROPIEDADES MODALES Y EPISTEMOLÓGICAS DE LOS ENUNCIADOS DE IDENTIDAD

A partir de estas ideas, Kripke desafía las creencias tradicionales acerca de las propiedades modales y epistemológicas de los enunciados de identidad. En la literatura corriente se advierte una marcada tendencia a intercambiar los términos «analítico», «necesario» y «*a priori*». Kripke juzga equivocada esta manera de proceder y la atribuye a la confusión de aspectos lingüísticos, metafísicos y epistemológicos. Recuerda, entonces, que una verdad analítica es aquella que depende de los significados de las palabras, de modo que la predicación de la analiticidad corresponde exclusivamente al plano lingüístico. En cuanto a las verdades *a priori*, subraya que son las que *pueden* conocerse independientemente de toda experiencia. De esta manera deja abierta la posibilidad de que haya verdades *a priori* que también puedan ser conocidas empíricamente. Y en lo que atañe al concepto de necesidad, comienza por señalar que algunos filósofos le dan un alcance epistemológico, en la medida en que hacen coincidir las verdades necesarias con las que se conocen *a priori*, y adelanta que usará el término como parte del vocabulario propio de la metafísica, y no de la epistemología. Debemos entender, en consecuencia, que otorga al concepto de necesidad una significación que se aparta tanto del sentido físico como del sentido lógico, tomado éste estrictamente, a menos que estemos dispuestos a suscribir la tesis de que la lógica se compromete por sí misma con la metafísica.

Es digno de notar, sin embargo, que pese a la importancia que le asigna a la distinción entre los planos lingüístico, epistemológico y metafísico, en algunas ocasiones Kripke transfiere ciertos conceptos de uno a otro terreno. Así es como utiliza las expresiones «necesario» y «contingente» —que de acuerdo con sus advertencias corresponden al plano metafísico— para calificar la verdad de los enunciados. La relación de identidad, por ejemplo, en tanto es precisamente una relación entre objetos, constituye una relación metafísica; y sostener que es necesaria equivale a decir que la propia relación es necesaria, independientemente de que se formulen o no enunciados que afirmen tal relación de identidad entre los objetos. Nubiola explica esta clase de desplazamientos desde el plano metafísico al de los enunciados, es decir, al ámbito lingüístico, vinculándolos con el concepto de esencia y la distinción entre modalidades *de dicto* y modalidades *de re*: un enunciado es necesario en la medida en que da cuenta de un hecho necesario, como ocurre con la identidad. La necesidad corresponde

primariamente a las cosas, es una modalidad *de re*, y de manera derivativa o secundaria llega a ser una propiedad de los enunciados que las describen (Nubiola, 1984, 248-249). Kripke no se ocupa extensamente de la distinción entre ambos tipos de modalidad; apela más bien al carácter intuitivo de las modalidades *de re*. Efectivamente, en las oportunidades en las que formaliza algunos de los enunciados que examina, especialmente cuando ejemplifica enunciados de identidad, utiliza el operador modal dentro del alcance de un cuantificador, lo cual evidencia que no se trata de modalidades *de dicto*. De todos modos, a partir de la introducción del concepto de designador rígido y la distinción entre los planos lingüístico, metafísico y epistemológico, Kripke arriba a conclusiones que contrastan con las creencias habituales. Concluye, en efecto, que hay enunciados necesarios cuya verdad se conoce *a posteriori* y enunciados contingentes de los que se sabe *a priori* que son verdaderos. Veamos, pues, en qué consiste su argumentación; consideremos, para ello, los distintos tipos de enunciados de identidad, conforme a la naturaleza de los términos que contienen:

a) Enunciados de identidad expresados por medio de dos nombres propios; por ejemplo:

(i) «Cicerón = Tulio».

Puesto que «Cicerón» y «Tulio» son designadores rígidos, se refieren al mismo individuo en todos los mundos posibles en los cuales dicho individuo existe. Como se supone, además, que la identidad es verdadera, el referente de ambos nombres coincide: la misma persona es denotada por los dos nombres y necesariamente esa persona es idéntica a sí misma; en consecuencia (i) es un enunciado necesario. Sin embargo, la verdad de (i) no se determina *a priori*, pues la intuición básica de que ambas expresiones son designadores rígidos no indica de ninguna manera la coincidencia de sus denotados; esto último se establece empíricamente.

b) Enunciados de identidad expresados por medio de un nombre propio y una descripción; por ejemplo:

(ii) Cicerón = «el autor de las *Catilinarias*».

Aquí la descripción no se comporta como un designador rígido, pues hay condiciones contrafácticas en las cuales el autor de las *Catilinarias* pudo no haber sido Cicerón y en cambio otros individuos pue-

den satisfacer la descripción en otros tantos mundos posibles. Por consiguiente, aunque en el mundo real (ii) es verdadera, de allí no se sigue que sea una verdad necesaria.

c) Enunciados de identidad expresados por medio de dos descripciones, por ejemplo:

(iii) «El inventor de los lentes bifocales = el primer director de Correos de los Estados Unidos».

En este caso se aplican consideraciones semejantes a la situación del ejemplo anterior, pues ninguna de las dos descripciones incluidas en (iii) es un designador rígido.

d) Enunciados de identidad expresados a través de un nombre propio y una descripción que denota el mismo individuo en todos los mundos posibles, por ejemplo:

(iv) 2 = «la raíz cuadrada positiva de 4».

En esta situación la descripción corresponde a lo que Kripke ha denominado un designador rígido *de facto* y dado que las dos expresiones que flanquean el símbolo de identidad en (iv) son designadores rígidos, (iv) es una verdad necesaria. En cuanto a la cuestión de si se trata de una verdad *a priori*, como estaríamos inclinados a creer, es conveniente tener en cuenta las observaciones de Kripke acerca de la posibilidad de que un enunciado *a priori* también sea conocido empíricamente. Como ya se ha señalado, conforme al uso que hace Kripke de los términos «*a priori*» y «*a posteriori*», estas expresiones no aparecen como excluyentes. Kripke cita el caso en el que se establece que cierto número es primo por medio de la utilización de una computadora. En esas circunstancias —afirma— lo que constituye una verdad *a priori*, en cuanto *podría* haberse alcanzado independientemente de la experiencia, surge de hecho *a posteriori*, porque nuestra fe en la información brindada por la máquina descansa en nuestra confianza acerca de su correcto funcionamiento, y esto a su vez supone la aceptación de la validez de ciertas leyes físicas. De modo que, si tal es el caso, la verdad necesaria acerca de que cierto número es primo es obtenida *a posteriori*. Si consideramos, pues, la posibilidad que se acaba de mencionar, (iv) representa un caso especial de enunciados necesarios *a posteriori*.

e) Enunciados de identidad expresados mediante dos descripciones que designan rígidamente, por ejemplo:

(v) La raíz cuadrada positiva de 4 = «el menor número primo».

A este caso se extienden fácilmente las consideraciones hechas con respecto a (iv).

Las alternativas consideradas a)-e) no agotan el inventario de las verdades necesarias *a posteriori*. Restan, por ejemplo, las que pueden denominarse «identidades teóricas» descubiertas por las ciencias fácticas. Por ejemplo, las hipótesis que identifican el fenómeno del calor con cierto movimiento molecular, o el agua con cierta composición química. En todos estos casos también estaríamos en presencia de verdades necesarias, de acuerdo con el criterio de Kripke, pues si el calor consiste en cierto comportamiento de las moléculas, lo mismo ha de suceder en cualquier mundo posible en el cual existan fenómenos calóricos, aunque puedan no estar acompañados de ciertas propiedades contingentes, como por ejemplo la aptitud para provocar determinadas sensaciones. En un mundo posible en el cual los seres humanos difiriesen de los hombres que habitan el mundo real por carecer de sensaciones térmicas, el calor no sería percibido; pero si nuestras teorías físicas son correctas, de todos modos el calor seguiría siendo lo que es: cierto movimiento molecular. La sensación correspondiente pudo usarse originariamente en nuestro mundo para fijar la referencia de la palabra «calor», pero este vocablo se incorpora al lenguaje como un designador rígido y aquello que sirvió para fijar su referencia —ya lo hemos visto— no constituye su significado. Posteriormente la ciencia descubre una propiedad esencial del calor y si realmente lo ha logrado (a través de una investigación empírica), ello obliga a reconocer una relación necesaria, más allá de la «ilusión de contingencia» asociada a la creencia de que la *sensación de calor* podría tener una causa distinta del movimiento molecular. Se trataría, entonces, de una relación necesaria, pero establecida *a posteriori*.

El ejemplo que acabamos de presentar no debe sugerir la conclusión de que Kripke está pensando en la necesidad meramente física. La relación que surge entre el referente de un designador rígido y las condiciones que lo acompañan en *todo mundo posible en que dicho referente existe* parece ser una relación más fuerte de lo que ordinariamente se entiende por necesidad física; ello explicaría por qué Kripke asume el compromiso de incorporar su concepto de necesidad en el ámbito de la metafísica.

Cierto es que, vistas las cosas de tal modo, no parece haber lugar para hablar de una necesidad física, diferente de la metafísica. A veces se dice, a propósito de la necesidad física, para oponerla a los hechos contingentes, que alude a las regularidades que valen en condiciones contrafácticas. Pero si asimilamos los mundos posibles de Kripke, como él mismo sugiere, con las condiciones contrafácticas, no queda claro cómo diferenciar la necesidad física de la necesidad metafísica.

Más sorprendente resulta ser la tesis de que existen verdades contingentes conocidas *a priori*. Kripke brinda como ejemplo el enunciado «La barra B mide un metro de longitud en t°» (donde «B» es el nombre del metro patrón de París y «t°» un instante determinado). Rechaza la posibilidad de que la verdad de dicho enunciado se establezca en virtud de una definición; una vez más aplica la distinción entre dar el significado de una expresión y fijar su referencia. Cuando se define el metro como unidad de medida indicando que es la longitud de la barra B, la definición sólo sirve para fijar la referencia, pero no para indicar el significado de la palabra «metro». De acuerdo con nuestras intuiciones lingüísticas, la expresión «un metro» es un designador rígido, porque el referente de «un metro» es invariante en todos los mundos posibles; mientras que «la longitud de B en t°» no designa rígidamente, dado que su referencia puede variar en circunstancias contrafácticas, la longitud misma es contingente. Así pues, aunque el enunciado sea considerado una definición de «metro», no establece propiamente ninguna sinonimia entre ambos designadores. Por otra parte, para cualquiera que haya fijado la barra B como patrón del sistema métrico, parecería que la verdad del enunciado «B mide un metro de longitud en t°» constituye un conocimiento *a priori*. Kripke concluye, entonces, que el enunciado en cuestión ejemplifica la posibilidad de que haya verdades contingentes que se conocen *a priori*.

8. CRÍTICAS DE LA TEORÍA DE KRIPKE

Las ideas de Kripke acerca del funcionamiento de los nombres, que coinciden con las propuestas de Putnam y Føllesdal, se han utilizado para encarar problemas filosóficos de muy distinta índole. Sin embargo, y aunque con escasa resonancia, varios filósofos han cuestionado las consecuencias de la nueva teoría de referencia. Salmon, por ejemplo, juzga plausible tanto la teoría referencial como las tesis esencialistas de Kripke, pero sostiene que esas tesis son independientes de la teoría de los designadores rígidos (Salmon, 1979, 703-724). El aná-

lisis de Salmon se centra en la aserción de Kripke acerca del carácter necesario de la relación que vincula un objeto con su origen. Kripke intenta mostrar que dado un objeto, por ejemplo, una mesa construida a partir de cierta pieza de madera, esa mesa no podría haberse originado en otro trozo de madera distinta. Por supuesto, si Kripke lograra probar esta afirmación, su alcance se extendería a cualquier clase de objetos, naturales o artificiales. Salmon sostiene que lo que Kripke quiere mostrar, en realidad, es que si resulta meramente posible que un objeto pueda originarse en un cierto trozo de materia, entonces ese mismo objeto no podría tener otro origen material. Ésta es, evidentemente, una tesis esencialista muy fuerte. Pero el interés de Salmon no es ponerla en duda, sino señalar que no se desprende de la teoría de la referencia de Kripke. Según Salmon, para arribar a su conclusión, Kripke supone también un principio según el cual si es posible que una mesa *x* se origine en un trozo de materia *y*, entonces necesariamente cualquier mesa originada en *y* es la misma mesa *x* y no otra. Ahora bien, esta clase de suposiciones es lógicamente independiente de la teoría de los designadores rígidos, pues esta teoría sustenta, fundamentalmente, que los nombres propios son designadores no connotativos, es decir, no equivalen a descripciones, y no dice nada, por sí misma, acerca de que dos objetos originados en una misma materia deben ser idénticos.

Por su parte, Quine ha señalado que el compromiso esencialista se asume en el momento en que se admite la cuantificación de la lógica modal. Pero entre los partidarios de la lógica modal hay grandes diferencias. Aunque algunos se declararían esencialistas sin sonrojarse, otros lo negarían de plano, como lo ha hecho Carnap. De manera que surgen varias cuestiones. En primer lugar, ¿hasta qué punto la lógica modal conlleva supuestos metafísicos de esa naturaleza? En segundo término, cabe preguntarse si la teoría referencial de los nombres es, como parece, una consecuencia natural de la aplicación de la lógica modal a la semántica. Por último, es necesario precisar en qué medida la teoría de Kripke incorpora tesis metafísicas que no son exigidas ni por la lógica modal ni por la descripción del funcionamiento de los nombres.

Con respecto al compromiso esencialista asumido por la cuantificación en la lógica modal y por la tesis de la designación rígida, Føllesdal también advierte la necesidad de distinguir grados de compromiso con el esencialismo (Føllesdal, 1986, 114-115). Estima que los argumentos de Quine están dirigidos a dos tipos diferentes de esencialismo. Por un lado —señala Føllesdal— encontramos una versión *débil* del esencialismo, que consiste en sostener que la necesidad es in-

herente a las cosas y no al lenguaje. La lógica modal cuantificada asume un esencialismo débil, en la medida en que afirma que un objeto por sí mismo, independientemente de que se lo nombre de una manera o de otra, posee ciertos rasgos de manera necesaria, mientras que las demás características las posee contingentemente. Este tipo de esencialismo es el que se encuentra en los sistemas de Carnap y de Lewis, y en consecuencia la actitud de estos autores —como lo ha advertido Quine— resulta incoherente, ya que por otra parte niegan todo sentido al esencialismo metafísico. En efecto, Carnap y Lewis conciben la necesidad como un concepto que puede explicarse a partir de la analiticidad, esto es, en relación con el lenguaje; pero esta creencia entra en conflicto con la intención de cuantificar en el interior de contextos modales. El esencialismo *fuerte*, en cambio, va más allá de la cuantificación de la lógica modal. Surge especialmente del intento de proporcionar criterios de identificación de un objeto a través de los distintos mundos posibles. Así, difiere también de la doctrina aristotélica, en tanto postula la existencia de esencias individuales, constituidas por un conjunto de propiedades que cada individuo posee exclusiva y necesariamente. Las críticas de Quine —señala Føllesdall— tienden a confundir ambos tipos de esencialismo, y es precisamente este indiscriminado rechazo lo que lo lleva a descartar la idea de la designación rígida.

Føllesdal juzga que las objeciones de Quine a la lógica modal cuantificada son demasiado desastrosas como para ser verdaderas; propone, entonces, una explicación del funcionamiento de los términos singulares que no apela a las esencias para identificar los objetos individuales. Quine había argumentado que a fin de evitar consecuencias paradójicas tales como que resulte verdadero el enunciado «Necesariamente (el número de los planetas es mayor que 7)», los partidarios de la lógica modal cuantificada deberían restringir el universo del discurso a entidades intensionales, objetos tales que cualesquiera dos propiedades que determinen unívocamente un objeto sean analíticamente equivalentes. En tal caso, todas las identidades verdaderas serían necesarias; más aún, todas las proposiciones verdaderas serían necesariamente verdaderas. Y, en vista de que toda proposición necesariamente verdadera es verdadera, la lógica modal sería absolutamente prescindible. Pero —señala Føllesdal— si bien el argumento de Quine estaba destinado a las modalidades lógicas, se extendería a cualquier intento de formular teorías adecuadas de la causación, los contrafácticos, el conocimiento, la creencia, etc., lo cual resultaría inaceptable.

Føllesdal intenta eludir estas consecuencias elaborando una teoría de la referencia independiente de toda suposición acerca de la exis-

tencia de esencias individuales. Traza una línea divisoria entre los términos generales y lo que denomina *nombres genuinos*, una suerte de designadores rígidos que se comportan de manera similar a las variables, por cuanto se refieren al mismo objeto a despecho de sus cambios o de la variación de nuestras opiniones o teorías acerca de ese objeto. En el lenguaje ordinario, los pronombres, la mayor parte de los nombres propios y algunas descripciones definidas actúan como nombres genuinos. Su propiedad distintiva es la transparencia referencial, es decir, la sustitubilidad *salva veritate* de los términos que tienen la misma referencia.

La lógica modal cuantificada combina la transparencia referencial con la opacidad extensional, que afecta los operadores modales e impide la sustitución irrestricta de unos términos generales por otros coextensionales en cualquier oración, pues no hay garantías de que se conserve el valor de verdad. De este modo, al hacerse compatibles la transparencia referencial con la opacidad extensional, se perfila una estrategia para enfrentar las destructivas consecuencias de las objeciones de Quine a la lógica modal. La opacidad extensional impide arribar a la conclusión de que todas las proposiciones verdaderas son necesariamente verdaderas, lo cual conduciría, como ya se ha sugerido, no sólo al derrumbe de la lógica modal sino también al desmoronamiento de las teorías propuestas para explicar la causación, el problema de los contrafácticos, etcétera.

La decisión de aceptar la lógica modal cuantificada y otros tipos de teorías no-extensionales semejantes equivale efectivamente a comprometerse con el esencialismo, como lo había predicado Quine, pero se trata del esencialismo débil. Tan débil, a juicio de Føllesdal, que no resultaría más objetable que la aplicación de operadores modales a fórmulas cerradas. En consecuencia, no se deberían oponer reparos a este compromiso, porque permanentemente hacemos uso de expresiones que figuran en contextos no extensionales cuando hablamos, por ejemplo, en términos contrafácticos o cuando asignamos probabilidades a un suceso. Esos contextos también exigen la opacidad extensional para que puedan ser cuantificados; de manera que, en la medida en que los utilizamos, todos somos esencialistas.

Føllesdal comparte, no obstante, el desagrado de Quine por el hecho de que se haya apelado a la versión fuerte del esencialismo para dar cuenta de la identificación de un objeto a través de los mundos posibles. Este tipo de esencialismo no sólo sería ajeno a la lógica modal: tampoco se necesita para asegurar la identidad del referente de un nombre genuino, tal como Føllesdal lo concibe. Aunque muchos partidarios de la lógica modal han creído ineludible recurrir a las esen-

cias individuales en el momento de dar una interpretación a la lógica modal, Føllesdal lo considera absolutamente innecesario. La preservación de la referencia, que es una propiedad fundamental de los nombres genuinos, no está vinculada con nociones modales ni implica ninguna clase de esencialismo. Que un término singular conserve la misma referencia a través de los mundos posibles es una consecuencia de su modo de referir, aun fuera de la lógica modal, así como sucede con las variables cuantificadas en cualquier tipo de contexto, sin que haga falta pensar en la lógica modal.

Puede observarse que, en general, hay una amplia coincidencia entre las ideas de Føllesdal y las de Kripke. Sin embargo Føllesdal no suscribe la explicación causal del surgimiento de los nombres propuesto por Kripke. Føllesdal reconoce que en algunos puntos su concepción está más cerca de la doctrina de Frege que de la de Kripke. Es así como juzga poco relevante el bautismo inicial de los objetos y se acerca a Frege en tanto sostiene que en su primera aproximación el nombre se refiere al objeto que mejor satisface las oraciones en las que dicho nombre figura y que son consideradas verdaderas por los hablantes. Pero se aparta de Frege en cuanto se niega a diferenciar un conjunto de oraciones que expresan el sentido del nombre y otro conjunto que expresa las creencias fácticas acerca del objeto nombrado. Føllesdal piensa que todas las oraciones contribuyen tanto para dotar al nombre de su referencia como para permitirnos descubrirla. Bajo la influencia de Quine, tiende a unificar los aspectos lingüísticos y las creencias fácticas, rechazando la distinción fregeana entre lenguaje y teoría. Subraya de este modo la interconexión de varios factores en el proceso a través del cual un término refiere y los hablantes encuentran su referencia. Entre estos factores se hallan el asentimiento o el rechazo de oraciones formuladas acerca del objeto referido, la ostensión, las acciones lingüísticas o no lingüísticas y nuestras teorías sobre la percepción y el razonamiento.

En la medida en que Føllesdal se aparta de la teoría causal de la referencia, la rigidez de los nombres aparece más como una meta regulativa que como un logro definitivo. Por ello es posible que un término singular sufra, a lo largo de su uso, un cambio inadvertido de su referencia, de manera que en sentido estricto su rigidez se vería afectada.

En su réplica a Føllesdal, Quine relativiza la distinción entre el esencialismo débil y el esencialismo fuerte (Quine, 1986, 114-115). Ambas versiones comparten la creencia de que un objeto, por sí mismo e independientemente de que se lo nombre de una u otra manera o no se lo nombre, posee algunos rasgos necesariamente y otros con-

tingentemente. Este supuesto, en el que se apoya tanto la lógica modal como la teoría de la designación rígida, es precisamente lo que Quine no está dispuesto a aceptar. Si bien reconoce que la teoría de la referencia propuesta por Føllesdal constituye una manera atractiva y conveniente de mirar el esencialismo, estima que sus ventajas alcanzan sólo a los contextos intensionales. Pero siempre ha pensado que la ciencia bien puede prescindir de esos contextos. No está de acuerdo, por ejemplo, en que la idea de causalidad nos obligue a ser esencialistas, ni siquiera, en el sentido débil. Quine sigue creyendo, pues, que la referencia ha de estar subordinada a la predicación, tal como sucede en los contextos extensionales, conforme a las ideas de Frege. Consecuente con su tesis de la relatividad ontológica concede más importancia a las oraciones que a la referencia de los términos. En tanto son las oraciones las que sirven de vehículo y piedra de toque de las teorías científicas, la referencia y la reificación no son objetivos de la ciencia. La noción de objeto resulta poco relevante, porque Quine da cuenta de los cambios que sufren las cosas concibiendo los cuerpos como extendidos tetradimensionalmente en el espaciotiempo, de modo que la preservación de la referencia no tiene un rol dominante. Aun en el caso de las variables, su comportamiento es menos decisivo que lo que se ha creído. Junto con la cuantificación componen un recurso para vincular oraciones observacionales y oraciones teóricas. Su contribución estructural al discurso consiste en ser el soporte de las funciones de verdad, lo que las liga. Es cierto que, en oraciones tales como «(Ex) (Fx . Gx, ...)» las variables hacen posible la referencia conjunta a objetos idénticos, pero ello no alcanza a implicar, a juicio de Quine, que la referencia prime sobre la predicación.

J. Almog también señala que la tesis de Kripke acerca del modo de referir de los nombres no permite deducir, por sí misma, ninguna propiedad modal de los enunciados (Almog, 1986, 210-242). Se trata de problemas diferentes y la manera en que uno de ellos se resuelva no obliga a adoptar determinada solución para el otro, como tampoco obliga a pronunciarse al respecto. Almog ilustra esta situación comparando la teoría de Kripke con la de Mill. Ambos coinciden en sostener que los nombres propios carecen de cualquier connotación descriptiva, pero sólo Kripke asocia esta condición con el hecho de que un enunciado exprese una verdad necesaria. Y ello se debe a que, en contraste con la manera en que había procedido Mill, aborda la investigación de los recursos designativos del lenguaje prestando atención previamente a los *designata* y a sus propiedades modales. Así, la rigidez que atribuye a los nombres ya nace de una orientación modal, porque lo que explicaría que «Nixon» sea un designador rígido, mien-

tras que «El presidente de los Estados Unidos en 1970» no lo es, radica en el hecho de que ser Nixon es una propiedad esencial de Nixon, en tanto que ser presidente de los Estados Unidos es una propiedad accidental.

Una de las manifestaciones de la adopción de la orientación modal es el reconocimiento de que algunas descripciones se comportan como designadores rígidos, en paridad con los nombres propios. Son aquellas que aluden a condiciones que resultan modalmente estables. Para establecer que una descripción es un designador rígido hace falta observar las características del objeto que la satisface más bien que algún aspecto del lenguaje; y para determinar si un designador rígido es fuerte o no hay que averiguar qué sucede en los distintos mundos posibles. La modalidad en la que Kripke piensa es de carácter metafísico, más que estrictamente lógica, como lo testimonia el uso de un test intuitivo para determinar, por ejemplo, si Nixon podría carecer de la propiedad de ser la descendencia de las Gametas G. Parece claro que en este caso no se trata de la mera posibilidad lógica —tal como ocurre con las descripciones de estado propuestas por Carnap— sino de un tipo de modalidad que está vinculada con la esencia del individuo.

Lo que Almog destaca, entonces, es que la lógica modal no conduce forzosamente a la tesis de los designadores rígidos, así como el hecho de reconocer que los nombres propios designan sin describir no implica nada acerca de las propiedades modales de los enunciados en los cuales figuran. Es importante notar que niega también que la lógica modal deba asumir cualquier clase de esencialismo. Esta opinión parece contraponerse a lo que sostienen Quine y Føllesdall. El nudo de la cuestión está en la interpretación de los sistemas modales. Quine observa que la manera de evitar consecuencias indeseables es restringir el universo del discurso a entidades intensionales de modo que las modalidades se mantendrían en el plano estrictamente lógico. Pero si la limitamos de semejante modo, la lógica modal pierde toda utilidad. Føllesdall, por su parte, considera imprescindibles las referencias a otros tipos de modalidades. En todo caso —y esto es lo que sugiere Almog— la lógica modal pura no tiene por qué pronunciarse acerca de cuáles entidades satisfacen sus fórmulas, y tampoco incluye indicaciones sobre qué clase de expresiones del lenguaje natural corresponden a sus símbolos. Podemos concluir que los tres filósofos, Quine, Føllesdall y Almog, coincidirían en la apreciación de que la lógica modal, en tanto teoría estrictamente lógica, no implica formalmente ninguna clase de esencialismo; pero cualquier uso relevante que se haga de ella no puede evitar comprometerse con el esencialismo en mayor o menor medida.

A. Gallois es otro de los autores que consideran que la rigidez de la designación no conduce directamente a las consecuencias que Kripke encuentra en ella (Gallois, 1996, 57-76). Gallois piensa que la expresión «designador rígido» es ambigua, ya que cabe distinguir dos tipos de designadores rígidos, a saber, los designadores rígidos *restrictos* y los *irrestrictos*. Sostiene que los nombres propios, tal como se utilizan ordinariamente, corresponden sólo a una de las dos clases mencionadas. Y concluye que la afirmación de una identidad en la que figuran designadores rígidos podría resultar contingentemente verdadera; en caso contrario, si la identidad es necesaria, se trata de una afirmación trivial.

A fin de ilustrar sus observaciones, Gallois propone el siguiente ejemplo. Supongamos que el barco Mary está construido con las tablas de un conjunto C, y que éstas son gradualmente reemplazadas por tablas de un conjunto disyunto C_1. Cuando el reemplazo ha sido total, el barco pasa a llamarse «Alice». En ese caso, podríamos decir que en el mundo real, W,

$$\text{Mary} = \text{Alice}.$$

En tanto, en otro mundo posible, W_1, se construyen dos barcos Mary_1 y Alice_1, compuestos, respectivamente, por las tablas C y C_1. Está claro que en W_1 Mary_1 es distinto de Alice_1. Pero, por otra parte, Mary es idéntico a Mary_1 y Alice es idéntico a Alice_1. Sin embargo, mientras en W es verdad que Mary = Alice, en W_1 esta identidad es falsa. Dicho de otro modo, Mary y Alice son idénticos en el mundo real, pero no en todo mundo posible, puesto que no lo son en W_1. En consecuencia, o bien la tesis de la designación rígida es falsa o bien «Mary» y «Alice» no son designadores rígidos.

Para resolver esta cuestión, Gallois descarta varias alternativas y en contraste con Kripke defiende la tesis de que Mary puede ser contingentemente idéntico a Alice, aun cuando «Mary» y «Alice» sean designadores rígidos. Para ello formula una condición necesaria y suficiente que caracteriza los designadores rígidos, a la que denomina *condición de la designación rígida* (RDC):

> RDC: (*d*) (*x*) (*W*) {*d* designa rígidamente a *x* en W ≡ (y) (W') [*d* (tal como es usado en *W*) designa a *y* en W'] ⊃ en W es verdad que $x = y$}.

Esta condición establece requisitos un tanto diferentes de los que Kripke señala como características de un designador rígido. Sin em-

bargo, Gallois piensa que se trata de un criterio plausible, especialmente porque resulta extensionalmente adecuado. Esto es, todas las expresiones normalmente consideradas designadores rígidos, y solamente ellas, cumplen con la RDC. Así, por ejemplo, puede mostrarse rápidamente que un nombre propio cumple esta condición. «Franklin», por caso, es un nombre que designa rígidamente a su referente, y en consecuencia, según la tesis de Kripke, designará en todos los mundos posibles a un individuo que es idéntico a Franklin en el mundo real. «Franklin» no puede designar a ningún otro individuo, por ejemplo Luis XIV, en otro mundo posible, porque Luis XIV no es idéntico a Franklin en el mundo real, de manera que «Franklin» también resulta ser un designador rígido conforme a la RDC. Recíprocamente, una descripción definida, tal como «el inventor de los bifocales», no es un designador rígido, de acuerdo con la RDC, porque esa descripción podría referirse, en algún mundo posible, a otro individuo distinto del que efectivamente inventó los bifocales, digamos Luis XIV, que no es idéntico a Franklin en el mundo real.

Veamos ahora cómo la adopción de la RDC conduce a la conclusión de que puede haber identidades contingentes formuladas en términos de designadores rígidos. Según una hipótesis previa,

A) en W: Mary = Alice

Por otra parte, se supone que «Mary» designa rígidamente a Mary en W y también designa a $Mary_1$ en W_1, pues como «Mary» es rígido, sólo puede designar en W_1 a un individuo que en W es idéntico a Mary, de acuerdo con la RDC, de modo que

B) en W: Mary = $Mary_1$

Por un razonamiento similar llegamos también a

C) en W: Alice = $Alice_1$

Aplicando la transitividad a las identidades A) y B) se obtiene

D) en W: $Mary_1$ = Alice

Y nuevamente por transitividad, esta vez en *C*) y *D*), se arriba a

E) en W: $Mary_1$ = $Alice_1$

Ahora bien, *E*) significa que en W hay un solo barco que corresponde a los que en W_1 son dos barcos distintos, de manera que E) no nos permitiría inferir

$$F) \text{ en } W_1: \text{Mary}_1 = \text{Alice}_1$$

La proposición *F*) sería, según nuestros supuestos, falsa. Sin embargo esta conclusión sería inevitable si sostenemos que «Mary» y «Alice» son designadores rígidos y entendemos por designador rígido una expresión «*a*» tal que si «*a*» designa rígidamente a *x*, entonces todo *designatum* de «*a*» debe ser idéntico a *x* no sólo en el mundo real sino también en todo mundo posible. Gallois propone llamar «designadores rígidos *irrestrictos*» a las expresiones que cumplen con esta condición. Pero hay otra opción: la RDC brinda la caracterización de los designadores rígidos *restringidos*, que obliga a que los *designata* de un designador rígido deban ser idénticos en el mundo real y puedan no serlo en otros mundos posibles, a la manera como sucede en el caso de «Mary» y «Alice». De este modo se evitaría llegar a la consecuencia *F*). En síntesis, quedaría justificada la alternativa que resuelve el problema del ejemplo propuesto por Gallois sosteniendo que «Mary» y «Alice» son designadores rígidos y sin embargo la identidad de Mary y Alice es contingente, dado que esto es posible cuando se trata de designadores rígidos restringidos. No queda excluida, de todos modos, la alternativa de insistir en que «Mary» y «Alice» son designadores rígidos irrestrictos, de manera que el enunciado «Mary = Alice» sigue expresando una identidad necesariamente verdadera. Pero, en ese caso, se estaría dando por sentado lo que se pretende probar.

Por otro camino, L. Roberts llega también a la conclusión de que los argumentos de Kripke no son definitivos (Roberts, 1985, 563-572). Roberts sostiene que cuando los nombres propios figuran en un enunciado de identidad cumplen simultáneamente dos papeles: por un lado, mantienen su referencia ordinaria, y por otro, expresan una aserción sobre modos de especificar la referencia. Kripke insiste en que cuando usamos los nombres propios a nuestra manera, manteniendo su referencia ordinaria, los enunciados de identidad verdaderos son necesarios. Roberts replica que el hecho de usar los términos «a nuestra manera» no impide que los enunciados tengan un componente metalingüístico. No le resultan convincentes las razones que esgrime Kripke para probar que el enunciado «Hesperus podría no haber sido Phosphorus» sólo podría ser verdadero si tuviera un sentido epistémico. Cree que interpretarlo así es incurrir en una argumenta-

ción circular, porque se presupone que un enunciado no es metalingüístico si no alude a maneras de especificar la referencia.

A propósito de cómo los enunciados de identidad podrían llevar a cabo la doble función de mantener la referencia ordinaria de los términos y a la vez incluir un componente metalingüístico o una indicación sobre modos de especificar la referencia, Roberts sostiene que la situación es similar a la que se puede encontrar, por ejemplo, en el caso de un niño que pregunta: «¿Puedo tomar una de esas fresas?», y recibe de su padre la contestación «Son cerezas, pero puedes tomar una».

Veamos ahora sus observaciones acerca de la otra línea de argumentación brindada por Kripke para probar el carácter necesario de la identidad, que incluye los siguientes pasos:

a) (x) (y) [(x = y) ⊃ (Fx ⊃ Fy)]
b) (x) Necesariamente (x = x)
c) (x) (y) {(x = y) ⊃ [Necesariamente (x = x) ⊃ Necesariamente (x = y)]}
d) (x) (y) [(x = y) ⊃ Necesariamente (x = y)]

En cuanto a la corrección formal, Roberts no presenta ninguna objeción. Pero lo que le merece reservas es el alcance que se pretende otorgar a *d*). Roberts cree que solamente permitiría concluir que un enunciado de identidad verdadero cuyas expresiones referenciales no tienen ningún contenido informativo, y en el cual sólo son relevantes las extensiones de los términos, es necesariamente verdadero. Pero Kripke se ha sentido autorizado a concluir que los enunciados de identidad verdaderos son necesarios cuando están expresados por medio de nombres propios. Como llega a esa conclusión fundándose en *a*), aplica una ley de sustitutividad que podría formularse así:

S) Los términos correferenciales son mutuamente sustituibles *salva veritate*.

Sin embargo, *S*) no es una fiel interpretación de *a*) por las siguientes razones. En primer lugar, *a*) es un enunciado expresado en el lenguaje objeto o modo material, mientras que *S*) está expresado en el metalenguaje o modo formal. En segundo lugar, en *a*) aparecen variables, que carecen de contenido informativo, mientras que *S*) alude a términos en general y no a términos sin contenido. Por último, *a*) está formulado en el cálculo de predicados, que es veritativo funcional (*truth functional*), mientras que *S*) no sufre esa restricción. La traducción lícita de *a*) sería, pues, la siguiente:

S') Si dos términos sin contenido se refieren a la misma cosa, toda función proposicional veritativo funcional que sea verdadera cuando es completada por uno de esos términos será verdadera cuando sea completada por el otro.

Esta interpretación de la consecuencia expresada en el último paso de la prueba formal *a*)-*d*) es mucho más débil que *S*) y no afecta, por supuesto, a la deducción misma. Pero muestra, si el análisis de Roberts es correcto, que el argumento formal elaborado por Kripke no demuestra que los enunciados de identidad expresados por medio de nombres propios deban ser necesariamente verdaderos; queda abierta la posibilidad de que sean contingentes.

9. CONCLUSIONES

Agregaré ahora mis propias observaciones críticas acerca de la tesis de Kripke[2]. En primer lugar, es curioso que Kripke desplace el concepto de necesidad desde la esfera de la metafísica a la esfera de los enunciados al mismo tiempo que censura prácticas similares. Nubiola trata de justificarlo sugiriendo que la necesidad debe entenderse primariamente como una propiedad ontológica, de manera que los enunciados llegan a ser necesarios secundaria o derivativamente, en la medida en que dan cuenta de una necesidad real. De acuerdo con esta propuesta, un enunciado de identidad verdadero expresado por medio de designadores rígidos resulta necesario porque describe un hecho necesario: la autoidentidad de un objeto. Cabe preguntarse entonces qué hechos describen los enunciados de identidad que contienen designadores no rígidos. Estaríamos tentados a decir que se refieren a identidades contingentes, pero Kripke se resiste a admitir que la identidad, en tanto algo que le sucede a las cosas, pueda ser contingente:

> Si «*a*» y «*b*» son designadores rígidos, se sigue que «*a* = *b*», si es verdadera, es una verdad necesaria. Si «*a*» y «*b*» no son designadores rígidos, no se sigue tal conclusión acerca del enunciado «*a* = *b*» (aunque los objetos designados por «*a*» y «*b*» serán necesariamente idénticos» (Kripke, 1981).

Esta tesis nos lleva a pensar que los enunciados de identidad que emplean designadores no rígidos no aseveran, hablando estrictamen-

2. Los temas expuestos en el presente trabajo están desarrollados con mayor amplitud en Gaeta, 1997.

te, ninguna identidad. Se trata de una extraña consecuencia para un análisis como el que emprende Kripke, que presume de ser menos artificioso que las explicaciones de sus oponentes. Pues, probablemente, la intuición de los hablantes —a la que Kripke suele recurrir— indicaría que todo enunciado que se pueda simbolizar con la forma: «*a* = *b*» asevera una identidad, independientemente de qué clase de designadores sean «*a*» y «*b*».

Es oportuno interrogarse también acerca de la función que cumplen los enunciados de identidad. El examen de las situaciones en los que se utilizan puede descubrirnos algunas de sus propiedades principales. Cuando se repite el mismo término a ambos lados del signo de identidad, ya sea que se trate de un nombre propio o de una descripción, su valor informativo es nulo y resulta difícil imaginar una ocasión en la que tengamos necesidad de decir algo así. Lo que comúnmente está en juego cada vez que se formula un enunciado de identidad no es la cuestión metafísica acerca de si una cosa podría ser distinta de sí misma. Predicar la identidad es útil en la medida en que se afirma que ciertas propiedades coinciden en un mismo objeto, y esta circunstancia es patente cuando se utilizan descripciones. Los nombres propios tienen algunas diferencias con las descripciones, es cierto, pero si no se asociaran de algún modo con propiedades de los objetos tampoco se comprendería cuál es su papel. El proceso por el cual aprendemos a usar un nombre propio no se inicia al habernos topado con el nombre «suelto». Generalmente, los nombres aparecen en un cierto contexto. Difícilmente cualquiera de nosotros pueda recordar la primera vez que oyó hablar de Cicerón, pero seguramente su nombre fue introducido en una oración que le atribuía alguna propiedad o algún acto. Posteriormente fuimos conociendo más características suyas; algunas las hemos olvidado y otras bien pudimos reemplazarlas porque cambiaron nuestras creencias. Pero eso no significa que cada vez que pronunciamos o escuchamos el nombre de Cicerón podamos dejar de pensar en alguna de sus múltiples descripciones. Si así fuera no sé de qué serviría enterarse en algún momento de que Cicerón y Tulio fueron la misma persona. Desde la perspectiva de Kripke, se replicará que las propiedades del objeto que se asocian con el nombre solamente sirven para fijar su referencia; pero me parece que una respuesta de este tenor no es más que un recurso para disimular la necesidad de tener en cuenta tales propiedades. Aun en el caso de los pronombres y las variables, que en principio carecen de toda connotación, debe observarse que sólo prestan un servicio útil cuando se las emplea conjuntamente con algún predicado. Imaginemos, por ejemplo, un texto en el cual el autor se refiere a alguien siempre a través

de un pronombre. Si la primera oración dice «Él echó café en la taza» y la siguiente afirma «Él agregó leche al café», cuando leemos la segunda oración normalmente identificamos su sujeto con el de la anterior, y así sucesivamente; esto es, vinculamos mentalmente el pronombre con una serie de descripciones. No encuentro motivos para negar que de este modo el pronombre adquiere una suerte de contenido o sentido, claro está que de carácter absolutamente contextual y provisorio. Algo similar puede decirse con respecto al uso de las variables. En relación con la manera en que la teoría de las descripciones de Russell soluciona ciertas paradojas, Kripke brinda la siguiente paráfrasis:

> Hay un objeto *x* tal que inventó los bifocales, y como cuestión de hecho contingente, hay un objeto *y*, tal que *y* es el primer director general de Correos de los Estados Unidos y, finalmente, es necesario que *x* es *y* (*ibid.*).

Es interesante observar, de paso, que Kripke antepone el operador de necesidad —que no aparece en la formulación original de Russell— a la cláusula que formula la identidad entre *x* e *y*, lo cual está en consonancia con lo que hemos señalado acerca de que, conforme a su teoría, los auténticos enunciados de identidad son los que expresan identidades necesarias. Pero el ejemplo sirve especialmente para mostrar que la utilidad de la inclusión de la cláusula de identidad está subordinada a las propiedades que el resto de la fórmula predica de *x* y de *y*. Lo interesante no es saber simplemente que *x* es igual a *y* o que la identidad es necesaria, sino que el inventor de los bifocales y el primer director de Correos son la misma persona, que ambos atributos en este caso coinciden.

Como han subrayado otros autores, el análisis de Kripke está condicionado por su adhesión inicial a un fuerte esencialismo. No es mi intención incursionar en una discusión metafísica acerca de la verdad de los postulados de esta doctrina. Pero en la medida en que constituyen más bien el punto de partida que la conclusión de sus argumentos, no creo que aporten elementos decisivos para dirimir la cuestión. Me limitaré a esbozar, muy esquemáticamente, una explicación alternativa de las cuestiones consideradas por Kripke.

Comencemos por el concepto de necesidad. Es indiscutible que la idea de que ciertos hechos suceden forzosamente, algunos son meramente posibles y otros son imposibles, forma parte de nuestro sistema prefilosófico de creencias. Tal parece que no podríamos aventurar ningún conocimiento si no presupusiéramos el uso de nociones mo-

dales. Pero, a despecho de la naturalidad con la que las empleamos, han resultado sumamente complejas, cuando no directamente paradójicas. El uso corriente de los términos modales cedió lugar tempranamente a una inquietante proliferación de acepciones filosóficas. La necesidad, a partir de la cual pueden definirse la posibilidad y la contingencia, ha merecido múltiples especificaciones: absoluta, relativa, ontológica, lógica, ideal, conceptual, analítica, natural, etcétera.

La desconfianza que han despertado los conceptos modales no es, pues, injustificada. Resulta dificultoso explicar en qué consiste o aun simplemente aludir a la necesidad de manera neutral, es decir, sin asumir compromisos filosóficos con hipótesis más generales. Valga como ejemplo que en la presentación informal que formulamos algunas líneas más arriba se habla de *hechos* de los que se predica la necesidad o la imposibilidad, con la consiguiente adhesión, aunque no sea más que por razones expositivas, a determinada ontología.

Pero quizá puedan aislarse, al menos, algunos tipos de necesidad cuyas implicaciones merezcan consenso en la comunidad filosófica. A primera vista, la que cuenta con mayores posibilidades es la necesidad lógica. Más allá de las diferencias que mantienen con respecto a otras clases de necesidad, muchos filósofos dan por cierto que las verdades lógicas tienen carácter necesario. Sin embargo, algunos autores han puesto en duda tal creencia. Edward Zalta —que caracteriza las verdades lógicas como aquellos enunciados que resultan verdaderos cualquiera sea la interpretación de sus constantes no lógicas y define la verdad necesaria en términos de enunciados que son verdaderos en todos los mundos posibles— proporciona ejemplos de verdades lógicas que no son necesarias (Zalta, 1985). Uno de ellos consiste en proponer, de acuerdo con estos criterios, la introducción de un operador de actualidad o realidad, representado por el signo $\mathcal{A}$ que se antepone a cualquier fórmula de un sistema formalizado y se define de modo tal que «$\mathcal{A}\phi$» es verdadera en un mundo posible si y sólo si ϕ es verdadera en el mundo real. Así, todas las instanciaciones de una fórmula esquema «$\mathcal{A}\gamma \rightarrow \gamma$» son verdades lógicas, ya que bajo cualquier interpretación el antecedente implica materialmente el consecuente. Pero bajo una interpretación donde γ es verdadera en el mundo real y falsa en algún otro mundo posible aparece una instanciación de la fórmula esquema que no es una verdad necesaria.

Podría pensarse en la conveniencia de forzar las definiciones de manera que todas las verdades lógicas permanecieran en la extensión de las verdades necesarias. Pero este camino no conduce muy lejos. En primer lugar, difícilmente permitiría elucidar el concepto de necesidad. (Observemos, de paso, que la definición de verdad necesaria

que propone Zalta tampoco constituye una elucidación satisfactoria, porque utiliza la noción de mundo posible, que ya incluye un concepto modal). Además cuáles sean las verdades lógicas depende de la lógica que se decida aceptar. Y, en relación con esto último, algunos autores —Mill y Quine, por ejemplo— se han negado a otorgar a las leyes lógicas un papel demasiado diferente del que le corresponde a las proposiciones de las ciencias fácticas.

Si se siguiera esta última vía, aún podría atribuirse carácter necesario a las leyes, pero se trataría de la denominada necesidad natural o física. Este concepto no parece haber tenido mejor suerte que las modalidades lógicas: muchos autores comulgan todavía con los escépticos argumentos de Hume. Se ha intentado esclarecer el concepto de necesidad natural vinculándolo con la presencia de condicionales contrafácticos, pero los problemas que generan están lejos de haber sido resueltos. De todos modos la argumentación pertinente conserva cierto aire de circularidad: se ha propuesto diferenciar las leyes naturales de las meras generalizaciones accidentales por el hecho de que las primeras tienen validez contrafáctica; pero también se propone admitir la verdad de un condicional contrafáctico en tanto esté apoyado en alguna ley.

Las dificultades mencionadas refuerzan la sospecha de que el intento de fundar la idea de necesidad en términos menos controvertibles está condenado al fracaso y justifican la adopción de una actitud escéptica al respecto. Pero, como ya he sugerido, es muy difícil imaginar cómo nos arreglaríamos si pretendiéramos prescindir de los conceptos modales.

Atentos a la aparentemente insustituible función de tales conceptos —recordemos las palabras de Føllesdall «todos somos esencialistas»— muchos han optado, siguiendo a Aristóteles, por asumir la necesidad como una característica metafísica inevitable. Sin embargo, existen alternativas menos drásticas. El reconocimiento de las ventajas que proporciona el poder disponer de las nociones modales no implica que deba considerárselas reflejos de propiedades trascendentes. Podríamos dar cuenta de ellas entendiéndolas como componentes de la red de mecanismos, conceptos y creencias que componen nuestra manera de representarnos la realidad —lo que podríamos llamar nuestro esquema conceptual— más que como parte de la realidad misma. El tratamiento que puede darse a la necesidad natural ilustra lo que pretendo decir. Si atribuimos a una hipótesis científica el carácter de ley, seguramente lo hacemos porque la consideramos suficientemente fundada dentro de nuestro sistema de creencias. En tal caso, afirmar que los hechos que la hipótesis describe son necesarios

no tiene por qué ir más allá que la fe que el propio sistema de creencias inspire. La utilización de los operadores modales, desde esta perspectiva, constituye principalmente un indicador del papel más o menos privilegiado y firme que asignamos a determinadas proposiciones.

La interpretación que estoy bosquejando será rechazada sin duda, por quienes están convencidos de que la necesidad radica en las cosas mismas. Pero, como se ha indicado a lo largo de esta exposición, esa actitud parte de la adhesión a ciertos postulados metafísicos que no todos estamos obligados a compartir. Se ha invocado a la intuición en apoyo de tales presupuestos —Kripke llega a decir que quienes discrepan con ciertas ideas suyas tienen las intuiciones invertidas—, pero esta presunta facultad humana ha quedado bastante desacreditada después que algunos viejos axiomas por mucho tiempo considerados como evidentes fueron abandonados.

Podrá objetarse también que el enfoque epistémico que propongo relativiza el concepto de necesidad hasta el punto de desvirtuarlo completamente. Desde que las concepciones justificacionistas del conocimiento científico perdieron vigencia y el ideal de la certeza absoluta pasó a ser considerado una quimera, todas las leyes de la ciencia resultarían, en este sentido, contingentes. Lo más coherente parecería, entonces, dejar de emplear los términos modales, como podrían sugerir los que adoptan una firme posición escéptica. Pero creo que pueden evitarse ambos extremos. Negar o afirmar la existencia de las modalidades *de re* no son las únicas opciones. Tiene sentido predicar el carácter necesario de ciertos enunciados conforme al papel que jueguen en el sistema de nuestras creencias, y así no tendríamos que abandonar el uso de algunos recursos que forman parte de nuestro esquema conceptual.

A partir de esta propuesta, podría establecerse una distinción entre la necesidad lógica y la natural, si es que estamos dispuestos a reconocer alguna diferencia entre el papel que cumplen las proposiciones que consideramos verdaderas por razones de orden lógico y las que requieren el concurso de la experiencia. Un examen más minucioso permitiría elaborar clasificaciones más ajustadas. Así, por ejemplo, dentro del dominio de lo que en un sentido amplio hemos denominado necesidad lógica, podríamos precisar las diferencias que subsisten entre las proposiciones analíticas y las que se definen como verdades lógicas estrictas.

La adopción de un enfoque epistemológico permite resolver adecuadamente el problema del *status* modal de los enunciados de identidad. No es forzoso asignar a las leyes correspondientes un alcance metafísico, como si fueran completamente independientes del pen-

samiento. Bastaría advertir que nuestro modo de organizar la experiencia supone la admisión de las leyes de identidad, sin asumir el compromiso de atribuirles ningún carácter trascendente, y concederles al mismo tiempo la posición relativa de mayor solidez dentro del conjunto de nuestras creencias. Esta manera de considerarlas permite seguir aplicando a los enunciados de identidad la distinción entre verdades necesarias y contingentes. Así, el enunciado «Franklin = Franklin» —a despecho de la escasez de oportunidades en las cuales pueda necesitarse formular identidades triviales— resultará necesario. En cuanto a enunciados tales como «Hesperus = Phosphorus» o «Cicerón = Tulio», bien pueden ser considerados contingentes, en la medida en que estamos dispuestos a reconocer que los nombres no están privados de cierto componente descriptivo. Esto último no quiere decir, sin embargo, que se identifiquen lisa y llanamente con las descripciones. Hemos visto que la relación entre un nombre y las descripciones asociadas con él es un tanto particular y permite cierta flexibilidad respecto de cuáles sean las descripciones correspondientes. Además, la presencia de dos nombres diferentes flanqueando el signo de identidad constituye un modo indirecto de indicar su equivalencia lingüística. Aunque no se trata de una implicación formal, autoriza a inferir inmediatamente que ambos nombres se usan para referirse al mismo individuo.

Por otra parte, las identidades que se expresan mediante descripciones pueden ser necesarias o contingentes, según el caso. Son necesarias si las descripciones aluden a propiedades que, de acuerdo con la organización de nuestro sistema de creencias, no pueden dejar de corresponder al objeto. Así, «9 = el cuadrado de 3» es una proposición necesaria, mientras que «9 = el número de los planetas» no lo es.

Desde la perspectiva que hemos adoptado, se facilita la conservación de la correspondencia del carácter modal de las proposiciones con su *status* gnoseológico. Las verdades que cumplen una función rectora especial en la organización del sistema, como las leyes lógicas o matemáticas, integran el dominio del conocimiento *a priori*. Las proposiciones para cuya determinación es menester —de manera directa o indirecta— el concurso de la experiencia constituyen creencias *a posteriori*, aun cuando el grado de certeza que las acompaña permita considerarlas como descripciones de hechos regidos por la necesidad natural.

Estas consecuencias chocan, evidentemente, con las que Kripke extrae de sus propios supuestos. El contraste no se debe solamente al rechazo del esencialismo metafísico postulado por Kripke y sus tesis semánticas. Obedece, además, a que me resultan aun menos convin-

centes sus ideas acerca del conocimiento. Creo que los argumentos con los que trata de probar que existen tanto verdades necesarias conocidas *a posteriori* como proposiciones contingentes conocidas *a priori* son defectuosos.

Es cierto que si se aceptan sus tesis acerca del carácter necesario de enunciados tales como «Hesperus = Phosphorus» o «agua = H_2O», y dado que no caben dudas de que su verdad sólo puede establecerse a partir de la investigación empírica, quedaría demostrada la existencia de proposiciones necesarias *a posteriori*. Pero Kripke trata de apoyar esta conclusión argumentando que la circunstancia de que algo pueda conocerse *a priori* no excluye que se conozca *a posteriori*. Y ejemplifica esta situación proponiendo el caso en el que se determina un resultado matemático mediante el uso de una computadora. Pues bien, creo que quienes sostienen la distinción entre ambas clases de conocimiento rechazarían semejantes ejemplos. Si, como sugiere Kripke, confiar en lo que la computadora indica es asumir creencias fácticas sobre su funcionamiento, no se diría que a través del artefacto *conocemos* la verdad del resultado matemático. Al menos, no con el alcance que se pretende dar a la expresión cuando se habla de conocimiento *a priori*. Si así fuera, simplemente no habría lugar para este modo de conocimiento: la resolución de las operaciones aritméticas requiere muy a menudo el empleo de lápiz y papel o algún recurso semejante, fuera de las computadoras; pero cuando alguien sostiene que hay conocimientos *a priori* lo hace porque considera que todos los aspectos sensibles, aunque estén presentes forzosamente, son irrelevantes para fundarlos. El razonamiento de Kripke parece tan descaminado como tratar de refutar a Kant recordándole que basta con contar guijarros para averiguar cuánto suman 7 más 5.

Más sorprendente es el ejemplo de la longitud del metro patrón, aducido para probar que existen proposiciones contingentes conocidas *a priori*. Kripke dice que «sabemos de alguna manera *a priori*» que la barra de París mide un metro, pero que la afirmación es contingente, pues en un instante cualquiera su extensión puede ser mayor o menor. Tengo la impresión de que el argumento es engañoso, porque si se toma la barra en cuestión como un patrón en el sentido estricto, un metro es —por definición— la longitud de la barra en toda circunstancia, mientras no cambiemos de patrón. En ese caso, la proposición se sabe verdadera *a priori*, pero como no podría ser falsa, también es necesaria. Si estamos inclinados a tomar en cuenta las posibles variaciones de la barra con respecto a alguna longitud ideal, en cambio, no sabemos nunca cuánto mide exactamente —a menos que contemos con otro patrón para compararla con él—, de manera que

no podemos asegurar ni *a priori* ni *a posteriori* que mide exactamente un metro.

Los agudos argumentos de Kripke, en síntesis, no son tan contundentes como para obligarnos a suscribir sus impactantes consecuencias. Podemos asignar un papel a los conceptos modales y a la noción de identidad sin asumir fuertes compromisos esencialistas. Podemos seguir pensando que a los enunciados de identidad necesarios les corresponde un conocimiento *a priori* y que sobre la base del conocimiento empírico pueden establecerse identidades contingentes, ya sea que se expresen por medio de nombres propios o descripciones. Como lo muestran los propios trabajos de Kripke, la consideración de los enunciados de identidad y el examen del papel que juegan los conceptos modales no pueden soslayar, finalmente, la relevancia de la perspectiva epistémica. Lejos de conducir a confusiones, esta perspectiva permite brindar una elucidación adecuada tanto de la identidad y de la necesidad cuanto de sus mutuas relaciones. Las ideas que al respecto hemos esbozado constituyen una base sobre la cual puede afirmarse la conveniencia de conservar un lugar para los enunciados que expresan modalidades e identidades sin necesidad de pagar el precio de la adhesión a la metafísica esencialista.

BIBLIOGRAFÍA

Almog, J. (1986), «Naming whithout Necessity»: *The Journal of Philosophy*, vol. LXXXIII, n.º 4.

Carnap, R. (1965), «La antigua y la nueva lógica», en A. J. Ayer (comp.), *El positivismo lógico*, FCE, México.

Føllesdal, D. (1986), «Essentialism and reference», en Lewis Edwin Hahn y Paul Arthur Schilpp (eds.), *The Philosophy of* W. V. *Quine* (Library of Living Philosophers), Open Court, La Salle, Ill.

Gaeta, R. (1997), *Lenguaje, identidad y necesidad*, CBC-Eudeba, Buenos Aires.

Gallois, A. (1996), «Rigid Designation and the Contingency of Identity»: *Mind*, 1996.

Humphreys, P. y Fetzer, J. [comps.] (1998), *The New Theory of Reference*, Kluwer Academic Publishers, Dordrecht/Boston/London.

Kneale, W. y M. (1992), *El desarrollo de la lógica*, Tecnos, Madrid.

Kripke, S. (1981), *Naming and Necessity*, Basil Blackwell, Oxford.

Nubiola, J. (1984), *El compromiso esencialista de la lógica modal*, Ediciones Universidad de Navarra, Pamplona.

Pescador, J. H. S. (1986), *Principios de Filosofía del Lenguaje*, Alianza, Madrid.

Quine, W. O. (1963), «Referencia y modalidad», en W. O. Quine, *Desde un punto de vista lógico*, Ariel, Barcelona.

Quine, W. O. (1986), «Reply to Dagfinn Føllesdal», en Lewis Edwin Hahn y Paul Arthur Schilpp [eds.], *The Philosophy of W. V. Quine* (Library of Living Philosophers), Open Court, La Salle, Ill.

Quine, W. O. (1965), *The ways of Paradox and other essays,* Random House, New York.

Roberts, L. (1985), «Problems about Material and Formal Modes in the Necessity of Identity»: *The Journal of Philosophy*, vol. 82, n.º 10.

Salmon, N. (1979), «How not to derive essentialism from the theory of reference»: *The Journal of Philosophy*, vol. LXXXVI, n.º 12.

Searle, J. (1972), «Proper names and descriptions, en P. Edwards [comp.], *The Encyclopedia of Philosophy*, Mcmillan and Free Press, New York.

Simpson, Th. M. (1975), *Formas lógicas, realidad y significado*, Eudeba, Buenos Aires.

Stroll, A. (1972), «Identity», en P. Edwards [comp.], *The Encyclopedia of Philosophy*, Mcmillan and Free Press, New York.

Tomás de Aquino, *Sum. Theol.*, I, 28, 1, ad. 2.

Torretti, R. (1980), *Kant*, Charcas, Buenos Aires.

von Wrigth, G. H. (1951), *An Essay on Modal Logic*, North Holland Publishing, Amsterdam.

Zalta, E. (1985), «Logical and Analytical Truth that are not Necessary»: *The Journal of Philosophy*, vol. LXXXV, 2.

EL CONCEPTO TARSKIANO DE VERDAD

Alberto Moretti

1. INTRODUCCIÓN

La noción de verdad frecuentemente juega un papel destacado en la consideración de varios problemas importantes. Tanto si esa consideración tiene pretensiones más o menos filosóficas, como cuando se realiza en el terreno práctico del «sentido común». Se sostiene, por ejemplo: (*a*) si una oración no es verdadera, lo que dice no transmite conocimiento; (*b*) la verdad de las creencias explica muchos de nuestros éxitos al intentar satisfacer nuestros deseos (en particular, da cuenta del éxito práctico de las teorías científicas); (*c*) el que creamos verdaderas nuestras creencias motiva muchas de nuestras acciones (en particular, nuestra búsqueda de teorías científicas); (*d*) la significatividad de las oraciones (las «informativas» cuando menos) depende de las condiciones en que son verdaderas o falsas; (*e*) la idea de verdad hace posible tomar inferencialmente en cuenta proposiciones de estructura desconocida y hace posible generalizaciones sobre proposiciones de estructura común; (*f*) las reglas de inferencia han de aceptarse debido a que conservan la verdad en el paso de premisas a conclusiones; (*g*) la caracterización de las afirmaciones morales y estéticas ha de involucrar una decisión sobre la posibilidad de su verdad o falsedad. Esta presencia de la noción de verdad en la gnoseología, la explicación de la acción, la semántica y la fundamentación de la lógica muestra que existen, *prima facie*, buenas razones para intentar su aclaración. Los procesos cognitivos procuran creencias verdaderas y éstas son valiosas para la acción. Cierto es que en cada uno de esos ámbitos cabe intentar proceder sin recurrir a la idea de verdad. De donde, quizá, debamos admitir que en ninguno de ellos es necesaria.

Pero esto dice poco en contra del esfuerzo por analizarla. Mientras no contemos, para esos problemas, con soluciones tanto o más plausibles que las que se basan en la noción de verdad, habrá motivos para creer que vale la pena procurar su esclarecimiento (al menos en relación con algunas de aquellas preocupaciones que no hayan sido satisfechas por medios alternativos).

Algunos pensarán que la ubicuidad y centralidad del concepto de verdad indica fuertemente que se trata de una noción básica, cuya intelección no podrá mejorarse a través de ningún conjunto de ideas pretendidamente más fundamentales. Aun así, el afianzamiento de esta posición no se logrará sin haber examinado los principales intentos por alcanzar una aclaración satisfactoria. Y, en cualquier caso, ese examen contribuirá a la comprensión del concepto mostrando algunas relaciones de las que participa. Sea con optimismo o sin él, parece, pues, inevitable plantearse el antiguo problema de la verdad.

Preliminarmente conviene hacer dos observaciones. La primera concierne a la aparente distinción entre determinar el significado o uso del predicado veritativo, por un lado y, por otro lado, proveer un análisis del concepto de verdad o de la naturaleza de la verdad. No prejuzguemos sobre la índole de esta diferencia, cuya discusión nos embarcaría en el examen del nexo entre lenguaje y pensamiento (examen que, por otra parte, tampoco podría alejarse demasiado de la cuestión de la verdad), pero admitamos que, *prima facie*, no puede tratarse de una diferencia tan radical como para no creer que la reflexión dirigida especialmente hacia una de esas cuestiones arrojará luz sobre aspectos importantes de la respuesta a la otra. La segunda observación remite también a una diferencia: la que media entre la búsqueda del significado de un predicado (o la naturaleza de un concepto) y la búsqueda de los métodos eficaces para su empleo de hecho. A veces podrá tratarse de la misma empresa, pero, en general, se trata de objetivos distintos. Por ejemplo, aunque descubrir recientes huellas digitales en trozos de cerámica del siglo V sea un método eficaz para aplicar a alguien el predicado 'es culpable de haber roto un valioso jarrón', haber dejado esas huellas es algo claramente distinto de ser culpable de haber roto un valioso jarrón.

2. EL CONCEPTO PRETEÓRICO DE VERDAD

El uso normal del predicado veritativo expone, implícitamente, el núcleo mínimo del concepto de verdad tal como fue culturalmente forjado. Por otra parte, los problemas filosóficos acerca de la verdad sur-

gen a partir de este uso normal. De modo que si el objetivo es determinar el concepto de verdad que parece requerirse para la gnoseología o la teoría lógica, es razonable que la tarea inicial abordada sea la identificación de ese núcleo mínimo. Dicho de otro modo: si no hubiera un núcleo preteórico especificable, ¿qué razón habría para pensar que la definición o teoría que se propusiera respecto de la expresión 'es verdadera' tiene algo que ver con algún problema filosóficamente interesante e históricamente recurrente vinculado con ese predicado? De no haberlo, la caracterización que se hiciera resultaría una mera estipulación de significado o una velada propuesta para sustituir un predicado irremediablemente confuso ('es verdadera') por un nuevo predicado que inaugure un giro conveniente de la «conversación» filosófica. (No hay que subestimar el interés de ciertos giros conversacionales, pero tampoco olvidar que la transformación espiritual, o algún beneficioso cambio en el modo de pensar y hacer, pueden llegar por o en medio de cualquier camino, incluso el sembrado de lógica formal y añejas cuestiones semánticas.)

La noción de verdad que se formule teóricamente deberá, pues, involucrar, como mínimo, ese núcleo conceptual que pueda determinarse de modo independiente y que servirá como piedra de toque para juzgar la bondad de teorías alternativas.

Cualquier hablante normal de un lenguaje natural como el que permite este texto (cualquier participante de la trama conceptual asociada con estos lenguajes) que acepte una oración P aceptará, sin más, la oración «P es verdadera», y viceversa. Es prudente notar que esta constatación no implica una respuesta clara a lo siguiente: quien no acepta P ¿acepta su negación?, ¿acepta siquiera que P es falsa? También es prudente desatender por ahora esta cuestión. Tarski ha insistido (y muchos concordado) en que en el hecho citado se resume el meollo del concepto de verdad, tal como aparece en el discurso normal del que deriva el llamado problema de la verdad. Más exactamente, ha sostenido que las instancias de sustitución del siguiente esquema (el esquema T):

X es verdadera si y sólo si P

(donde 'X' ocupa el lugar de un nombre de una oración que dice lo mismo que la oración cuyo lugar ocupa 'P') concentran el contenido mínimo del concepto de verdad o, por lo menos, caracterizan el funcionamiento habitual del predicado veritativo (Tarski, 1956, 155, 187; 1944, § 4; 1969, 64; ver además Hochberg, 1992). Se dice que, en cada caso, la oración que sustituye a 'P' enuncia las condiciones ve-

ritativas de la oración nombrada por el nombre que sustituya a 'X' (esto es, las condiciones cuyo cumplimiento o las circunstancias cuya ocurrencia se requiere y es suficiente para hacer verdadera a esta oración). El esquema T establece una asociación entre la oración mencionada con 'X' y el uso de la oración que sustituya a 'P'. En un sentido mínimo de «representación», establece que la oración X y lo extralingüístico que pudiera asociarse con ese uso de P se representan mutuamente (así como suelen hacerlo, por ejemplo, las verdades lógicas y las pruebas). ¿Deberemos considerar que la tesis anterior, que llamaremos (T), sobre que el sentido mínimo de 'es verdadera' compromete con la aceptación de los casos (paradigmáticos) del esquema T, es una tesis empírica acerca de su uso natural y de su contenido conceptual? Podría pensarse, en efecto (cf. Tarski, 1969, 63), que expresa el aspecto descriptivo concurrente para forjar la elucidación del predicado. Sin embargo, ¿admitimos la refutabilidad de (T)? Si algunos hablantes lingüísticamente competentes rehusaran aceptar las equivalencias indicadas por el esquema T (en particular, algunos casos «típicos»), ¿mostraría esto que (T) es falsa, que aquel esquema no forma parte del concepto de verdad, o sería más una razón para creer que esos hablantes no asocian nuestro concepto de verdad con su uso del predicado veritativo? Parece mejor pensar que (T) es una tesis *a priori*, constitutiva del significado de 'es verdadera' y del contenido de la noción de verdad. Nosotros (los que nos involucramos con problemas históricamente legados como los citados en el § 1) llamamos concepto de verdad al que se manifiesta en el uso lingüístico característico de un predicado como nuestro predicado 'es verdadera'. Que cierta comunidad lingüística *tenga* (explícitamente) el concepto de verdad, es refutable. Pero no es refutable que un concepto, para ser un concepto de verdad, deba satisfacer algo como (T). No lo es porque ésa, según parece, es la característica mínima constitutiva del significado de ese predicado. Así es que cualquier aclaración de esta idea tendrá que incluir esta tesis como componente esencial. Claro está, siempre que la tesis no implique que, a pesar de su prestigio, la noción de verdad es insalvablemente confusa.

Concedamos que, muchas veces, la aceptación de una oración X es un fenómeno relevante para la discusión acerca de la noción de verdad. Cuando eso ocurre cabe decir, en general, que se acepta X con la pretensión de que sea verdadera, esto es, *con la pretensión de que* X es verdadera, pero no *sobre la base de que* X es verdadera. Cuando alguien dice X y creemos que lo que ella dice casi siempre es verdadero, es por *estas* creencias y no meramente por la creencia, derivada de éstas, de que X es verdadera, por lo que solemos aceptar X.

No habríamos podido creer que X es verdadera sin que alguien —tal vez esa persona en quien confiamos— hubiese tenido creencias justificatorias de X independientes del concepto de verdad. Podemos aceptar X antes de tener el concepto de verdad (y ser entonces incapaces de aceptar que X es verdadera), pero no podemos aceptar que X es verdadera sin aceptar X. Aceptamos que X es verdadera *sobre la base de que* X es aceptada (con pretensiones de verdad). Si tuviésemos motivos para afirmar X, tendríamos motivos para afirmar «X es verdadera». Ésos serían motivos para afirmar que X es verdadera, *por ser motivos para* afirmar X.

Las condiciones veritativas de X (cuando se dan o cuando son afirmables) determinan que X es verdadera (o que es afirmable con pretensiones de verdad). Si el usuario normal del predicado veritativo es consultado: ¿por qué 'P' es verdadera?, responderá: porque P. Si, en lugar de aquello, se le preguntase ¿por qué P?, no se aceptaría, más que como énfasis aprobatorio, una respuesta como: porque 'P' es verdadera. La pregunta era por la justificación o por los rasgos del mundo. Las respuestas podrán ser: porque Q, o: porque el mundo es así.

Que 'Marte gira' es verdadera (en el sentido corriente de 'verdadera') no es algo convencional, arbitrario, decisional, definicional. Así pensaba, entre otros, Leibniz (véase *Diálogo sobre la conexión entre las cosas y las palabras*, 1677). Nuestra voluntad puede cambiar que 'Marte' refiera a Marte o 'gira' se aplique a lo que gira. Pero, fijado eso, está fuera de nuestro control (conceptual) que 'Marte gira' sea verdadera (o falsa). Eso lo *determina* el mundo. O, para usar un conjuro más reciente, los hechos semánticos relativos a la verdad sobrevienen a partir de hechos no semánticos. Este rasgo: que el darse de las condiciones veritativas de 'Marte gira' determina la verdad de 'Marte gira', también es constitutivo de nuestro concepto de verdad (recuérdese, por ejemplo, Aristóteles, *Categorías*, 4 a,b).

Esta observación plantea un segundo criterio de adecuación para cualquier aclaración del concepto ingenuo de verdad. La legitimación del siguiente esquema (F) (donde la relación aludida es asimétrica):

X es verdadera en virtud de que P

Al decir «X es verdadera» usualmente queremos decir que X tiene una especial relación con el mundo, determinada por el significado de las expresiones suboracionales y por cierto aspecto, en general parcial, del mundo, conectado con el significado de esas expresiones suboracionales. Que eso queremos decir se pone de manifiesto cuan-

do, supuesto que 'P' sea una oración y admitido que la entendemos, también admitimos *a priori* que

'P' es verdadera *en virtud de que* P

(y no al revés). O, quizá,

'P' es verdadera *porque* P

O, quizá,

Si 'P' es verdadera entonces ('P' es verdadera *porque* P)

Si es así, la aclaración de 'es verdadera' deberá preservar la verdad de estos contextos. (Es evidente, en este artículo, que el uso de comillas a veces desatiende el rigor técnico con la excusa de aliviar la exposición.)

El uso del predicado veritativo establece una relación entre lo dicho por las oraciones 'P es verdadera' y 'P'. En cierto modo, los criterios (T) y (F) son maneras de desplegar el contenido de esa relación.

Las consideraciones antecedentes presuponen, es claro, que la expresión 'es verdadera' es un predicado genuino que alude a cierta propiedad predicable de oraciones. Estas presuposiciones, sin embargo, pueden cuestionarse. Algunos piensan que, pese a las apariencias, el significado profundo de 'es verdadera' muestra que no se trata de un predicado sino de un operador para construir variables de oraciones. Otros, por su parte, aceptan su carácter predicativo pero creen que las entidades de las cuales se predica no son oraciones sino proposiciones o, quizás, creencias o, aun, aseveraciones. No discutiremos aquí estos puntos (a tal efecto, véase Russell, 1912; Austin, 1950, Grover, 1992) confiando —en lo que al último respecta— en que las cuestiones centrales que se estudien, o bien sean esencialmente independientes de la elección que se efectúe, o bien determinen esa elección (bajo la idea de que el problema de esta elección no es independiente de los otros pero es, en todo caso, secundario respecto de éstos).

Además de las tesis (T) y (F), otro rasgo que parece preteóricamente dado, ya vimos, es el vínculo de la verdad con la capacidad cognitiva y la satisfacción de intereses prácticos. La forma más sencilla de asegurar la conexión entre la idea de verdad de una afirmación y nuestros recursos epistémicos, justificatorios de la aceptación de algunas afirmaciones y no de otras, consiste en definir la verdad en términos de esos procedimientos. Sosteniendo, por ejemplo, que una aseveración es verdadera precisamente cuando nos es posible ofrecer

(o por lo menos reconocer) una justificación suficiente para aceptarla. Esta garantía se verá, tal vez, en la existencia de algún proceso probatorio especial, o en la posibilidad de pasar la prueba de una discusión ideal, o en la coherencia con un conjunto amplio y asentado de creencias, o, quizás, en sus méritos como guía para la acción. Impresionados por el nexo entre verdad y éxito práctico, también podría intentarse definir la verdad directamente en términos de capacidad para promover acciones exitosas. Estos barruntos están en el origen de teorías de la verdad que han sido llamadas verificacionistas, consensualistas, coherentistas y pragmatistas (véanse Hempel, 1935; White, 1957; Ezorsky, 1967; Dummett, 1959; Hesse, 1980). La verdad de las oraciones está vinculada, en efecto, con los métodos para su aceptación y con las consecuencias prácticas de aceptarlas, pero también, suele creerse, con la representabilidad del mundo en general. El que aseverar algunas oraciones se prefiera a hacerlo con otras se explica, según esto, por la pretensión de representar el mundo (ya que tener éxito en esta empresa cuenta como la mejor explicación del éxito práctico derivado de aseverar algunas y no todas las oraciones; y fallar en ese intento explica la decepción de expectativas fundadas en creencias). En un sentido de «representar» más comprometido que el expresado por (T). En el sentido de copiar al mundo, reflejarlo o, por lo menos, señalar la manera como nos afecta. Un sentido más cercano a (F). Cuando se entienda que esta última intención representacional es prioritaria y conceptualmente excluyente, se verán, las anteriores, como teorías basadas en intentos desviados por transformar meros criterios útiles para la aplicación tentativa del predicado veritativo, en claves para el análisis de la naturaleza de la verdad (véase la observación final del § I). Éste es el punto de vista de las llamadas teorías correspondentistas de la verdad, según el cual las teorías aludidas antes contestan directamente a las preguntas ¿para qué queremos creencias verdaderas? o ¿por qué creemos verdaderas algunas creencias? e, indebidamente, toman sus respuestas como si también lo fueran para ¿qué es que una creencia sea verdadera? La réplica general a esta objeción será, claro, que nada hay de indebido en aquella aparente simplificación, si resulta que todo uso teórico respetable de la noción de verdad puede mantenerse cuando se la entiende de un modo que hace de su naturaleza algo fácilmente accesible a los seres humanos. Los correspondentistas sostendrán, por su parte, que si el análisis de un concepto no remite directamente a la relación entre el lenguaje y el mundo, entonces no se trata del concepto de verdad que aparece en el lenguaje común y en los problemas filosóficos históricamente dados. Y verán esas teorías no como análisis de una noción

tradicional (pero pretendidamente inútil y perjudicial) sino como propuestas de sustituirla (y sustituir con ella su red de problemas) por otra noción, quizás postmoderna y servicial, a tono con un cambio cultural más sabio o placentero.

3. TARSKI Y LA *ADAEQUATIO* CLÁSICA

En un sentido amplio, hasta sostener que no puede haber ninguna aclaración interesante del concepto de verdad cuenta como aclaración suya. Las interesantes, sin embargo, serán las que se ofrezcan como análisis filosófico del concepto. Las formas más notorias de estos análisis son la definición explícita y la formulación de principios que determinen el lugar de la noción estudiada en alguna trama conceptual pertinente. Estos principios, típicamente, pueden ser axiomas al modo de los sistemas matemáticos usuales o procedimientos de reformulación de contextos amplios (quizá oraciones) en los que aparezca el concepto examinado. En ambos casos, una cuestión formal obvia es la de la consistencia (de la definición o del conjunto de principios). El hecho de que este problema suele ser más arduo en el segundo caso fue una importante razón por la cual Tarski optó por la vía definicional.

En general, en una aclaración conceptual se adoptará una lista de nociones que se pondrán en relación con el concepto analizado. Ésta será una relación de prioridad, en el orden de la fundamentación, cuando el análisis tome la forma de una definición; o será un vínculo que no presuponga ninguna jerarquía entre las nociones relacionadas, cuando el análisis sea del otro tipo. En el primer caso se pretenderá que las ideas que formen el análisis sean más claras que el concepto estudiado; en el otro caso se buscará, en lo posible, que las nociones utilizadas no sean manifiestamente más huérfanas de inteligibilidad que la analizada. El análisis, además, sea estrictamente definicional o aparentemente más laxo, habrá de mantener los contenidos del concepto que preteóricamente surjan como motivando el interés por conservarlo, y deberá también decidir acerca del grado de precisión y la amplitud con que, ahora teóricamente, habrá de usarse.

La definición aparece, tradicionalmente, como caso paradigmático o ideal de aclaración conceptual. En ella se establece algún tipo de equivalencia entre nociones. Y hay varios tipos. Puede ofrecerse una identidad de sentido, o una equivalencia lógica, o una equivalencia necesaria pero de tipo distinto al de la necesidad lógica. O puede darse una mera equivalencia extensional, confiando en que el rótulo «de-

finición» le confiera algún grado metalingüístico de necesidad, aunque nada en su forma lo señale. Lo menos que ha de exigirse a una definición (y, aun, a cualquier caracterización) de un concepto, será que legitime su aplicación y no aplicación a los casos que preteóricamente aparezcan obvios. Una aclaración interesante sugerirá, además, motivos por los cuales explicar esas aplicaciones y no aplicaciones e insinuará ciertas maneras, y no otras, de extenderlas.

Veamos ahora la posibilidad de aclarar la noción de verdad teniendo en cuenta el paradigma definicional. Si se adopta el punto de vista correspondentista, las condiciones de adecuación de la definición, además de la corrección formal, parecen ser dos: garantizar la aceptabilidad de las instancias del esquema T y reflejar la presunta correspondencia entre las oraciones y el mundo. Hacia 1931, Tarski procuró llevar a cabo este proyecto (Tarski, 1956, 153; 1944, § 3; 1969, 63) en el marco de exigencias metodológicas cuya satisfacción permitiera incorporar la definición, y el concepto así definido, a cualquier examen de teorías científicas rigurosamente formuladas, tanto empíricas cuanto puramente formales. En particular, con una caracterización de este tipo, la noción resultaría adecuada para la evaluación de sistemas de lógica.

Hasta cierto punto, especialmente para quienes sólo se sienten cómodos en aguas extensionales, cada uno de los bicondicionales del tipo T, como

'La luna gira' es verdadera si y sólo si la luna gira

define parcialmente el predicado veritativo. En efecto, cada uno de ellos determina algunas condiciones en las que el predicado es aplicable, y señala también a qué oración sería aplicable si se dieran esas condiciones. Además, puede formularse un bicondicional como ése respecto de todas las oraciones del lenguaje (por lo menos de lenguajes que cuenten con recursos como para formar nombres propios de cada una de sus oraciones). La conjunción de todos ellos, pues, parece dar criterios suficientes como para establecer la extensión del predicado veritativo. En tanto, como se dijo, lo menos que cabe esperar de una definición es que permita determinar la extensión del *definiendum*, y dado que cada uno de los bicondicionales ofrece un criterio para incorporar, o no hacerlo, una oración dada a la extensión de 'es verdadera' y, en este sentido, formula una definición parcial, la conjunción de todos resulta, en cierta medida, *una* definición total de ese predicado. Tarski pensaba, efectivamente, que los bicondicionales T «explican de modo preciso, de acuerdo con el uso lingüístico, el sig-

nificado de frases de la forma 'x es una oración verdadera' que aparecen en ellos» (Tarski, 1956, 187). A partir de lo cual estipuló que si una definición, formalmente correcta, de un predicado de oraciones tiene entre sus consecuencias a todas las instancias del esquema T, *entonces es* una definición *adecuada* de la verdad. Es decir, implicar los casos de T basta para ser considerada *una* definición del predicado veritativo que explica su funcionamiento en el lenguaje. Por lo demás, opinaba, si el lenguaje incluyera sólo un número finito de oraciones, y fuesen nombrables, «la construcción de una definición correcta de verdad no presentaría dificultades. Para este fin sería suficiente completar el siguiente esquema:

> x es verdadera sss o bien $x=x_1$ y p_1, o bien $x=x_2$ y p_2, o ..., o bien $x=x_n$ y p_n

reemplazando los signos 'x_1', 'x_2', ... , 'x_n' por nombres de las oraciones del lenguaje, y los signos 'p_1', 'p_2', ... , 'p_n' por oraciones correspondientes que digan lo mismo que aquéllas (Tarski, 1956, 188). En los casos que más nos importan, el número de oraciones es infinito, de modo que este procedimiento no puede llevarse a cabo directamente. Sin embargo, el método que se emplee, y la forma final de la definición que se logre, parecen no tener otra restricción que la implicación de las oraciones T, ni otra pretensión que la alcanzable mediante un hipotético esquema, similar al empleado en el caso finito excepto por presentar una disyunción infinita en su término derecho.

Garantizar la verdad de los casos de T era un desiderátum porque estos casos se presentan como verdades indudables acerca del significado o uso del predicado veritativo. Garantizarlo, además, permite dar inmediata cuenta de varios rasgos fundamentales de ese uso. Una de las funciones características del predicado veritativo es posibilitar la referencia a oraciones cuya estructura detallada se ignora, a fin de suponerlas, rechazarlas, creerlas o incorporarlas a procesos inferenciales. Véase este diálogo. A: Irma dijo algo que implica que los dinosaurios no existieron; no se qué dijo exactamente, pero, puesto que los dinosaurios existieron, creo que eso es falso; B: Ella dijo que la Tierra tiene sólo cinco mil años; A: La Tierra tiene más de cinco mil años. Otra función esencial es la de permitir generalizaciones acerca de infinitas oraciones, como al sostener que todos los casos del esquema 'Si P entonces P' son verdaderos. Ambos rasgos son fácilmente justificables si aceptamos que la oración 'P es verdadera' es equivalente a 'P', cualquiera sea ésta. Es posible imaginar otros medios para cumplir estas funciones, por ejemplo mediante cuantificación sustitucio-

nal y variables oracionales, pero es dudoso que alguno de estos medios sea tan sencillo como el provisto por el predicado veritativo y tan claramente presente como éste en el lenguaje natural. De modo que la necesidad de garantizar los casos de T está plenamente justificada. Pero ¿eso basta para reflejar la intuición correspondentista?

La idea correspondentista tradicional, que se remonta a Platón y Aristóteles (por ejemplo, *Sofista* 261e5-263b12; *Metafísica* Γ 1011, Θ 1051; *Categorías* 4b, 14b) pasando por la escolástica y buena parte de la filosofía moderna, reúne tres componentes: la existencia de un nexo fuerte entre oraciones verdaderas y hechos específicos (el componente representacional), la dependencia de la predicación de verdad respecto de la existencia de hechos (el componente fundacional), y la naturaleza en general extramental o no epistémica de los hechos (el componente realista). Las diversas teorías que responden a esta idea difieren, a veces grandemente, en el modo en que interpretan estos rasgos generales. Un modelo típico discurre así: una oración X *es verdadera* si y sólo si existe un hecho Z (de índole, en general, extramental) tal que X corresponde a Z; y X *corresponde a* Z si y sólo si los componentes $e_1, ..., e_n$ de X refieren a los componentes $h_1, ..., h_n$ de Z, y la forma en que están relacionados $e_1, ..., e_n$ en X remite (o es idéntica) a la forma en que están relacionados $h_1, ..., h_n$ en Z; y, además, X es verdadera *porque* se corresponde con Z, un hecho existente. Así es que podría haber hechos inexpresables, pero no oraciones verdaderas sin hechos correspondientes. Eventualmente todos ellos podrían ser inexpresables, *i.e.* podría no existir ninguno de los hechos requeridos para verificar las oraciones de la mejor teoría acerca del mundo que podamos imaginar (posibilidad intolerable para antirrealistas y realistas moderados).

Una teoría que, tomando literalmente la sugerencia tarskiana, según la cual una lista como la dada más arriba es una definición adecuada para el caso finito, buscara *solamente* reproducir, para el caso infinito, el efecto de una lista semejante, tendría una forma como la siguiente: todos y nada más que los casos del esquema 'X es verdadera si y sólo si P' son los axiomas de la teoría de la verdad. Con buena voluntad se verá que esta teoría establece un nexo entre las oraciones, nombradas por los sustitutos de 'X', y el mundo, «nombrado» por los sustitutos de 'P'. Con algún misterio, se dirá que, aquí, no se expresa la existencia del hecho de que P sino que se la muestra. Lo cierto es que nada se dice sobre existencia de hechos, *status* ontológico del mundo o relación de fundamentación (el 'si y sólo si' es simétrico), aunque, enseguida veremos, hay quien piensa que el criterio (F) —que remite a ese nexo de fundamentación— es subsumible

bajo (T). Así, pues, en ningún sentido interesante esta teoría expresa suficientemente la idea tradicional de correspondencia. Si tal se deseaba, la teoría deberá modificarse, aunque quizá también se requiera algún cambio en la interpretación de aquella idea tradicional.

4. LA TEORÍA MÍNIMA

Llegados a este punto, algunos han optado por sostener que la única misión conceptual del predicado veritativo es esa que, ya vimos, puede cumplir con que su significado se agote en la legitimación de las equivalencias T (véase Horwich, 1990). Estos autores afirman, entonces, que una teoría-lista como la bosquejada en el párrafo anterior es, ni más ni menos, la teoría de la verdad. Y se esforzarán por mostrar que cualquier otro papel conceptual que la noción parezca cumplir —en la explicación del conocimiento o la acción— se puede justificar sin necesidad de otorgarle más contenido que el que le adjudica su parca teoría. Dirán que la idea correspondentista de la verdad, y el criterio (F) en particular, es una confusión proveniente de trasladar, a la noción de verdad, notas que dependen de su relación con otras nociones —como la de referencia— pero que no le son constitutivas (no pertenecen a «su esencia»). Este enfoque es heredero de la concepción redundantista, adjudicable a Ramsey y tal vez a Frege, se aproxima a la concepción descomilladora de Quine y es pariente cercano de la reciente teoría pro-oracional de Grover y Belnap (véanse Ramsey, 1927; Frege, 1918; Quine, 1970; Grover, 1992), todas las cuales integran el tronco de las que —por motivos evidentes— se conocen como concepciones deflacionarias de la verdad (deflacionarias o mórbidas, según los afectos de quien juzgue).

Una teoría mínima como la indicada exhibe la peculiaridad de no utilizar ninguna noción general sustantiva (*i.e.*, de contenido más que lógico-formal) en su desarrollo. Está claro, entonces, que quienes anhelaban un *análisis* eliminativista no la encontrarán satisfactoria. Pero aun quienes no buscaban tanto ¿por qué preferirán un «análisis» tan magro frente a otro que explícitamente relacionara la idea de verdad con otras, tales como referencia o significado, con las que —sin duda— tiene relación? Si no se pretende explicarla en términos de otras nociones más básicas, ¿por qué preferir axiomas que sólo usan el concepto de verdad a axiomas que emplean, además, otras nociones? Quizá se piense que los axiomas más «ricos», oficiando como definición implícita, sugieren erróneamente conexiones necesarias —constitutivas de significado— donde sólo hay nexos secundarios (*a poste-*

riori). Pero es difícil ver dónde está el error: ¿dónde trazar la demarcación entre las funciones de un concepto constitutivas de significado y aquellas que no lo son? En momentos así llega la invocación a la simplicidad. Pero la simplicidad es relativa a objetivos y, hasta ahora, ha sido siempre algo elusivo. ¿Cuál simplicidad: la de los axiomas, la de las explicaciones suministradas, o la de alguna combinación de ambas?; ¿en relación con la explicación de los fenómenos que dieron motivo a la teoría o tomando en cuenta también su contribución para dar cuenta de otros fenómenos?; ¿según el número de conceptos intervinientes o la estructura lógica que los combina?; ¿respecto de nuestra teoría actual sobre el mundo, o relativamente a una teoría ideal?

Una estrategia normalmente seguida cuando se trata de comprender algún asunto, consiste en buscar rasgos y vínculos «profundos»; esto es, condiciones que no aparecen en la superficie, en la descripción inicial del tema, pero que —por motivos vagamente discernibles— generan una explicación acompañada de un sentimiento de mayor satisfacción intelectual. El empleo de esta estrategia, unido a la creencia de que una de las funciones esenciales del concepto de verdad reside en aludir al papel explicativo de lo extralingüístico (extramental) en la justificación de la diferencia entre oraciones afirmables y no afirmables (para fines cognitivos), apoyada en la creencia preteórica asociada con el criterio (F), conduce a pensar que la teoría mínima antes bosquejada no es suficiente. Esa teoría no es más que el sucedáneo de una lista infinita de oraciones T, y una lista así sólo es una tosca ruta hacia la extensión de 'es verdadera', que no explica, que no ofrece *un* concepto explicativo aplicable a *todos* los casos de aplicación de 'es verdadera'. La teoría mínima asocia la aplicación del predicado veritativo con condiciones distintas (en general) para cada oración del lenguaje (ej.: 'Llueve', con que lloviese; 'Hoy es lunes', con que hoy fuese lunes). El problema principal que se le imputa a la lista no es que sea infinita (y que entonces no sea explicitable), sino que no es explicativa (en un sentido de 'explicación' —hay que reconocer— que no pasa de ser una desarticulada creencia habitual). El criterio (F) sugiere que hay una conexión explicativa entre las condiciones de verdad de X y la predicabilidad de 'es verdadera' a X. Lo que, a su vez, induce a pensar que ha de haber algún rasgo peculiar de X que la conecta con que en el mundo se dé P más que lo que la conecta con que se dé Q. Esos rasgos contribuyen a explicar por qué aceptamos X en lugar de su negación. El papel conceptual del predicado veritativo no parece agotarse, entonces, con el oficio de expediente práctico para el comercio lingüístico.

Pero Horwich ha sostenido que la teoría mínima es compatible con (F). En su opinión, es metodológicamente erróneo integrar (F) en el análisis de la verdad, pero, admite, (F) expresa un rasgo genuino de nuestro empleo de la noción y, en consecuencia, el análisis debe poder justificarlo. El argumento metodológico es probablemente más débil que el de la simplicidad —que ya lo es suficientemente— a menos que sea el mismo. El argumento que le permite hallarse «perfectamente cómodo con la idea de que cada verdad se hace verdadera por la existencia de un hecho correspondiente», es el que sigue. Las leyes naturales junto con las condiciones iniciales del universo explican (aceptemos que deductivamente implican) que la nieve es blanca. «La nieve es blanca» queda establecida. La teoría mínima afirma que «La nieve es blanca» es materialmente equivalente a «'La nieve es blanca' es verdadera». Deducimos, pues: «'La nieve es blanca' es verdadera». De esta manera la explicamos y justificamos que

(1) 'La nieve es blanca' es verdadera *porque* la nieve es blanca;
y también que
(2) El que 'La nieve es blanca' sea verdadera *se explica por* el ser blanca de la nieve.

Pero, *pace* Horwich, no toda deducción es una explicación. Ni todo paso deductivo es un paso explicativamente relevante. Apelando a las leyes y al estado inicial del mundo tenemos una explicación de la verdad de 'La nieve es blanca', pero ésa no es una explicación basada en que la nieve es blanca. Y es esto lo que (1), como caso de (F), dice. Supongamos que

'Dalmacia tiene riñones' es verdadera si y sólo si la nieve es blanca.

Las premisas que explican que la nieve sea blanca no explican la verdad de 'Dalmacia tiene riñones'. Claro que aquí la equivalencia en juego no nos es cognoscible *a priori*, como se pretende de los casos de (T). Pero el 'si y sólo si' pasaba por ser extensional (aun si cognoscible *a priori*), sin embargo ahora aparece con una carga intensional dada por su *status* epistémico, pero ¿de dónde, sino del sentido de 'es verdadera', proviene ese *status*? Las oraciones T son a priori precisamente *por el sentido de 'es verdadera'*. Eso (el tal sentido) hay que revelar para justificar la aprioricidad de las oraciones T. No basta adoptarlas diciendo que son *a priori*. La carga intensional requerida no es la de la sinonimia (porque la teoría mínima justificadamente rechaza esa tesis; piénsese al respecto en las oraciones con deícticos). Y

tampoco basta con adjudicarle carácter de instancia de una generalización empírica. Supongamos este caso:

Dalmacia tiene riñones si y sólo si Dalmacia tiene corazón.

Admitamos que tenemos una explicación del corazón de Dalmacia. Aun si esas mismas premisas explicaran la existencia de sus riñones, la equivalencia anterior no nos muestra que también el que posea corazón explica o determina el que posea riñones, ni viceversa.

Aquello que explica que la nieve es blanca es lo mismo que explica que 'La nieve es blanca' es verdadera. Pero esto no es explicar la verdad de 'La nieve es blanca' sobre la base de la blancura de la nieve. Si en la equivalencia entre «'La nieve es blanca' es verdadera» y «La nieve es blanca», que es uno de los axiomas de la teoría mínima, estaba ya el componente explicativo, entonces no era un caso de (T) sino de (F). Y si no estaba, entonces la teoría mínima no justifica (1) ni, en general, (F).

Otro problema acucia a la teoría mínima. Amenaza, es cierto, a cualquier teoría que implique los casos del esquema T, pero dada la índole de la concepción mínima, es a ésta a la que inquieta más. Sencillamente, el esquema T, irrestricto, lleva a contradicción. La gramática castellana usual admite que 'Esta oración es breve' y, análogamente, 'Esta oración es falsa' son oraciones castellanas. Sustituyendo en T, se tendrá que 'Esta oración es falsa' es verdadera si y sólo si esta oración es falsa. El contexto aclara que 'Esta oración es falsa' es la oración aludida en el lado derecho de esa instancia de T. De donde: 'Esta oración es falsa' es verdadera si y sólo si 'Esta oración es falsa' es falsa. Lo que conduce a afirmar, utilizando lógica clásica elemental, que existe una oración que es y no es verdadera. La contradicción puede obtenerse sin utilizar expresiones deícticas obvias y sin recurrir a observaciones contextuales del tipo empleado. Pero basta el caso dado para probar que, respecto de lenguajes naturales corrientes, no es cierto que puedan admitirse *todos* los casos del esquema T. De modo que si T expresa adecuadamente el sentido normal del predicado veritativo, tenemos motivo para que la noción definida por la teoría no sea idéntica a ese concepto ingenuo.

Si la teoría mínima hiciera de las oraciones (entendidas idealmente) las entidades sobre las cuales se predica la verdad, podría intentarse una resolución del problema estipulando que el predicado veritativo no puede formar parte del lenguaje al que pertenecen las oraciones a las que se aplica. Pero el mérito principal de la teoría mínima es su aparente sencillez, que se pretende ligada al hecho de que

respeta los usos lingüísticos normales, no contaminados teóricamente, y que pretende, tan sólo, recogerlos (sin adherirles rasgos «profundos»). Es posible que, pese a las apariencias en contrario, el lenguaje natural esté escindido en varios lenguajes que no comparten sus predicados veritativos (quizá como resultado de estar gobernado por un principio general de rechazo de contradicciones). O que la extensión de ese predicado derive de un proceso de aprendizaje controlado, en último término, por la capacidad de usar oraciones que no lo emplean (Kripke, 1975). Pero, como quiera que sea, si lo que parecía más obvio (el esquema T) es erróneo, cualquier restricción debe justificarse teóricamente. Y si la sencillez teórica ha de perderse de todos modos, ¿por qué insistir en un abordaje inicial conceptualmente restringido? Cabe imaginar que un planteo explicativo que procure organizar, en el marco de la teoría de la verdad, varias intuiciones concurrentes (una de las cuales fuese el esquema T), podría facilitar el hallazgo de razones atendibles para justificar las restricciones que se impongan sobre ese esquema.

5. EL MÉTODO Y LAS DEFINICIONES TARSKIANAS

Tarski fue el primero en intentar, de modo sistemático, la integración de la idea correspondentista tradicional de la verdad junto con la aceptación restringida del esquema T (Tarski, 1956, 153, 164; 1944, §§ 3, 9, 17; 1969, 63, 67). Expresándose a veces como si creyera que este esquema expone, con mayor claridad, el contenido de aquella idea. Sin embargo, ya se observó, la noción tradicional correspondentista-realista involucra más que lo ofrecido por (T). Por otro lado, también se ha señalado muchas veces —el mismo Tarski se encargó de hacerlo— que el esquema T es sólo el criterio para juzgar la adecuación de cualquier definición que se proponga. Cabe pensar, pues, que en la genuina definición se reflejen los rasgos centrales característicos de la idea tradicional (particularmente aquellos que no se reproducen en T).

En los círculos científico-filosóficos frecuentados por Tarski, el concepto de verdad parecía a la vez fundamental y demasiado oscuro como para integrarlo en el análisis riguroso del conocimiento y, en particular, para estudiar los fundamentos de la lógica y la matemática. Su incorporación teórica amenaza, sabían, con contradicciones. Por otra parte, la escuela hilbertiana desarrollaba con éxito un programa sintacticista de fundamentación. Y los resultados alcanzados por Gödel mostraban la posibilidad de desembarazarse de conceptos semánticos y, sin embargo, lograr la demostración de profundos teo-

remas, incluso acerca de los límites de aquel programa, cuya formulación preteórica apelaba a la idea de verdad. En este contexto Tarski se propuso mostrar que, respetando el sentido que la tradición filosófica le fue dando al predicado veritativo del lenguaje natural, es posible definir la extensión de un predicado razonablemente coextensivo con ése, y tal que su empleo teórico no conduce a inconsistencias y tiene suficiente contenido y claridad como para ser parte de la empresa fundacional.

La solución tarskiana de las paradojas que rodean al concepto de verdad depende de la tesis de que el predicado veritativo (y los demás predicados semánticos problemáticos) no forma parte del lenguaje a cuyas expresiones se aplica. De aquí se sigue la necesidad de establecer la teoría de la verdad, relativa a un lenguaje, utilizando otro (un metalenguaje) que permita referirse al que se estudia y que contenga el predicado veritativo correspondiente. La consecuente restricción del esquema T, impidiendo la existencia de casos donde se use el mismo predicado veritativo en ambos miembros del bicondicional, es suficiente para evitar la paradoja del mentiroso. Resta ver si puede hallarse alguna base para aquella tesis que, aunque se asiente en la idea de verdad, sea independiente de la necesidad de resolver las antinomias (algo análogo al principio del «círculo vicioso» russelliano para justificar su teoría de tipos). Deben mencionarse otras dos consecuencias importantes de este enfoque: el metalenguaje deberá tener mayor poder lógico-expresivo que el lenguaje-objeto, ya que, si tuviese el mismo, el predicado veritativo que definiera pertenecería también al lenguaje-objeto reinstalando la antinomia; asimismo, se advierte, no podrá existir una definición de 'verdadera' aplicable a todo lenguaje (cualquier definición deberá excluir a los lenguajes de mayor o igual poder lógico-expresivo que aquel en el que se formula).

Una decisión metodológica tácitamente adoptada por Tarski fue la de utilizar solamente recursos propios de un lenguaje extensional, a fin de producir la definición. Sin duda, el deseo de garantizar en grado máximo la inteligibilidad de la construcción motivó este requisito. Las nociones sustantivas o explicativas que se empleen y la lógica con que se las maneje deberían ser de mínima problematicidad. La lógica (clásica) de los lenguajes extensionales (o de los tramos extensionales de lenguajes cualesquiera) satisface este último punto. Además, Tarski ejemplificó su método para construir la definición, respecto de un lenguaje también extensional y que no lo obligó a utilizar ninguna noción sustantiva específica. Esto es, no necesitó emplear concepto semántico alguno, valiéndose tan sólo de los predicados de he-

cho presentes en el lenguaje estudiado, eventualmente enriquecidos con algunas ideas lógico-matemáticas (cuya comprensión, aparentemente, está garantizada).

Como vimos, las definiciones en que Tarski pensaba para lenguajes finitos son conceptualmente tan ricas como el esquema T, no más. Y la definición que buscaba para los casos de infinitas oraciones trataba de no alejarse conceptualmente (*i.e.*: en cuanto a los conceptos que se utilizacen) de una generalización infinita de aquella definición. Esto es una fuerte indicación de que, a pesar de su declarada intención de recuperar —con su definición— la idea tradicional de verdad, Tarski no pretendía ir más allá de lo expresado por T. Buscaba que el contenido de su definición fuese esencialmente el mismo que el de T. Otro indicio, en la misma dirección, lo da la ya señalada suficiencia del cumplimiento del criterio T, para la adecuación de la definición. No se trata, como podría pensarse, de una señal inequívoca, ya que el criterio lo es de la adecuación de *cualquier* definición de verdad, no necesariamente de una que pretendiera ser correspondentista. Al menos cabe pensarlo de este modo, aunque su autor lo haya formulado en el contexto de la búsqueda de una defínción de ese tipo especial. Un indicio más se citará después.

La vía hacia la definición se simplifica si se considera que el predicado veritativo es aplicable a las oraciones. Sin prejuzgar acerca de si éstos son los objetos originarios de la predicación de verdad; sin creer siquiera que ésta sea una cuestión interesante. Pero suponiendo que, en caso de no ser las oraciones sino otras entidades relacionadas con ellas, las pasibles de verdad, esas relaciones sean lo suficientemente estrechas como para que la simplificación realizada no impida trasladar los resultados logrados. Estas oraciones, por otra parte, no coinciden con meros grafismos o sonidos, sino que son portadoras de significatividad; están interpretadas (Tarski, 1956, 166, 167). Tienen un aspecto sintáctico dado por pautas generales que asemejan o correlacionan, gráfica o auditivamente, emisiones (de grafismos o sonidos) que cumplen un papel similar en el intercambio lingüístico. Debido al carácter de estos objetos, está claro que la extensión de 'es verdadera' podrá variar (y, en general, lo hará) con cada cambio de lenguaje. Así pues, será posible *prima facie* solucionar el problema de definir ese predicado *relativamente* a cada lenguaje particular, sin esperar una definición general válida para todo lenguaje, algo imposible, como vimos. Lo que cabe buscar es un *método* suficientemente general como para permitir construir la definición específica para cada uno de los miembros de una clase amplia e importante de lenguajes. Y, aunque no se ofrezca una caracterización rigurosa y gene-

ral del método, un gran paso se dará mostrando ejemplos de cómo procede en casos paradigmáticos. Éste es el enfoque de Tarski.

El problema planteado por los lenguajes interesantes es el de la infinitud potencial de sus oraciones. Una de las ideas básicas para solucionarlo es la tesis de la existencia de un número *finito* de *formas básicas* a las que se ajusta esa infinitud oracional. La clave determinante de esas formas son los signos lógicos. Porque ellos rigen las inferencias y éstas, a su vez, son el mecanismo lingüístico que restringe la distribución de valores veritativos entre las oraciones. Supongamos que un lenguaje L tuviese un número finito de oraciones elementales (Q_1, Q_2, ..., Q_n), pero contara también con un número finito de conectivos extensionales, digamos '~' y '&'. Tenemos, con L, un ejemplo sencillo de un lenguaje con infinitas oraciones, pero con sólo tres tipos básicos oracionales. En un caso así, la definición de verdad podrá darse, aproximadamente, en estos términos:

(1) Q_1 es verdadera sss q_1
Q_2 es verdadera sss q_2
....
Q_n es verdadera sss q_n
(2) ~ A es verdadera sss A no es verdadera
(3) A & B es verdadera sss A es verdadera y B es verdadera

[donde 'A' y 'B' son sustituibles, o bien por oraciones elementales (el tipo de 1), o bien por oraciones construidas a partir de otras mediante la aplicación de '~' o '&' (tipos 2 y 3)].

De donde resultará algo como (cf. § III):

X es verdadera sss o bien (X = Q_1 y q_1) o bien (X = Q_2 y q_2) o bien ..., o bien (X = Q_n y q_n) o bien (X = ~A y A no es verdadera) o bien (X = A&B y A es verdadera y B es verdadera).

Un problema más serio (porque lo presentan los lenguajes naturales) se plantea cuando también es infinito el número de oraciones elementales. Se logrará una solución, análoga a la anterior, si el lenguaje cuenta con sólo un número finito de generadores de oraciones elementales. Uno de esos mecanismos es la predicación. La cuantificación es otro. Ambos procesos dependen de la posibilidad de reconocer expresiones suboracionales de categorías distintas: términos singulares (nombres propios y descripciones definidas) y predicados. Sean esas expresiones las que fueren, si el lenguaje presenta solamente un número finito de ellas, entonces, con auxilio de proce-

sos como la predicación y la cuantificación, pueden construirse todas las otras.

Siguiendo toscamente una tradición fregeana, cabe considerar predicado al resultado de «horadar» una oración cualquiera mediante la extracción de algunas apariciones de sus términos singulares. Dado que el número de oraciones es infinito, y que con cualquier número finito de ellas se puede formar una oración conjuntiva, el lenguaje deberá contar con algún medio para indicar los infinitos predicados complejos que resultan de la aplicación de aquel procedimiento. Naturalmente, esta posibilidad no tiene consecuencias prácticas, pero la teoría sistemática debe tenerla en cuenta. La manera más sencilla de hacerlo es la postulación de la existencia, en el lenguaje, de un número (potencialmente) infinito de términos singulares especiales, las variables individuales. Éstas, a diferencia de los demás términos singulares, no se interpretan como refiriéndose a objetos específicos, sino que actúan como pronombres, indicando los lugares en los que cabe sustituirlas por nombres propios (o descripciones definidas si las hubiere). Su presencia, sabemos, es decisiva para la expresión inambigua de las oraciones cuantificacionales. Su sola presencia en la categoría de términos singulares, además, permite la generación de infinitas oraciones e infinitos predicados complejos (para un lenguaje que incluya algún conectivo binario) a partir de un número finito de predicados simples ('Fx', 'Fx*Fy', 'Fx*Fy*Fz', ...; '(Ex) Fx', '(Ex) (Ey) (Fx*Fy)', ...).

La clave del método tarskiano aplicado a lenguajes con infinitas oraciones elementales reside en la solución del problema de dar condiciones veritativas para todas ellas, habida cuenta de que, para estas oraciones, la determinación de su valor veritativo no es función del valor veritativo de oraciones componentes (que no tienen), sino, en todo caso, del valor referencial de sus expresiones suboracionales. Lo esencial del método a seguir puede mostrarse considerando un lenguaje sin nombres propios, con infinitas variables individuales y con un único predicado simple. Tarski usó un ejemplo de este tipo, que ahora se bosquejará (Tarski, 1956, § 2). Si el predicado —digamos 'I'— fuese binario y se interpretara como el concepto de inclusión entre conjuntos, y las variables fueran 'v_1', 'v_2', 'v_3', ..., entendiéndolas referidas a conjuntos, las expresiones como 'Iv_7v_3' serán llamadas predicados complejos, funciones proposicionales o cuasi-oraciones. Y expresiones como '(Ev_7) (Ev_3) Iv_7v_3' serán oraciones del lenguaje. La solución utilizada por Tarski consiste en construir un concepto más general que el de verdad, en el sentido de que sea aplicable tanto a oraciones como a cuasi-oraciones, a partir del cual definir éste (Tarski, 1956, § 3). El elegido fue el concepto de que los objetos o_1, ..., o_n

satisfacen el predicado n-ésimo φ. A fin de evitar la relativización a un número finito de objetos (que obligaría a contar con infinitos conceptos de satisfacción) conviene proceder con la siguiente noción, más general: *la secuencia infinita σ de objetos satisface el predicado φ.*

Las variables del lenguaje están ordenadas por los números naturales, y por tanto lo están las de cada cuasi-oración. Relativamente a cada secuencia de objetos, además, se estipula que la variable i-ésima refiere al objeto i-ésimo (denotado por 'σ_i') de la secuencia σ. Para el ejemplo elegido, los objetos son conjuntos. Así, se define:

σ satisface φ sss	o bien φ es $Iv_k v_l$	y σ_k está incluido en σ_l
	o bien φ es $\sim \psi$	y σ no satisface ψ
	o bien φ es $\psi \vee \xi$	y σ satisface ψ ó σ satisface ξ
	o bien φ es $(v_k)\ \psi$	y toda secuencia que difiere de σ a lo sumo en el k-ésimo lugar satisface ψ

Esta definición, de carácter recursivo, puede transformarse en otra, explícita, al precio de cuantificar sobre relaciones. Sólo mediante esta transformación el predicado 'σ satisface φ' es eliminable en todo contexto, pero lo crucial es advertir que, en ambos casos, el análisis (eliminativista o recursivo) no emplea más nociones que las presentes en el lenguaje del que se habla (el predicado metalingüístico 'está incluido en' «traduce» el predicado 'I' del lenguaje-objeto) junto con otras, tal vez ausentes en ese lenguaje, pero de carácter lógico-matemático-sintáctico ('no', 'o' y 'toda' también pertenecen al lenguaje-objeto, pero tal vez no 'difiere', 'secuencia' y otras, usadas para referirse a expresiones de ese lenguaje).

Las oraciones reaccionan de modo peculiar frente a la satisfacibilidad por medio de secuencias. Es claro que la posibilidad de que unas secuencias satisfagan, y otras no, una cuasi-oración, depende de la existencia de variables libres (no cuantificadas). Ya que podrá ocurrir que el objeto correlacionado con una variable libre, digamos v_k, respecto de cierta secuencia, sea uno que verifica el predicado, pero sea otro que no lo verifica, cuando se pase a una secuencia que difiera de la anterior en el lugar k-ésimo. Pero una oración no tiene variables libres, de modo que, o bien todas las secuencias la satisfacen o no lo hace ninguna. Y si alguna la satisface, entonces lo hacen todas. Llegamos así a la célebre definición:

X es verdadera sss toda secuencia satisface X

A costa de complicaciones técnicas en la definición de 'σ satisface φ', se puede obviar el recurso a secuencias infintas y proceder sobre la base de secuencias finitas, construidas eliminando de las infinitas todos los objetos que son irrelevantes para las variables de la oración o cuasi-oración. El procedimiento es igualmente legítimo desde el punto de vista conceptual, al menos según Tarski, y hasta tiene la ventaja de resultar más «natural» que el anterior (cf. § VI). Sólo por razones de simplificación puramente técnica, Tarski usa secuencias infinitas.

6. DEFINICIONES TARSKIANAS, REALISMO Y CORRESPONDENCIA

Veamos ahora si en la definición de verdad podemos encontrar reflejados los rasgos tradicionalmente asociados con la idea correspondentista. Parece claro que Tarski no buscó deliberadamente ese reflejo. Su construcción fue dictada por consideraciones dirigidas al único propósito de lograr que la definición implicara todos los casos del esquema T. Sin embargo, independientemente de sus intenciones, merece examen la posibilidad de que sus definiciones expresen, mejor que T, o que la teoría mínima (§ IV), el contenido tradicional de la verdad como correspondencia.

La definición tarskiana parece explicar la verdad en términos de una relación, la de satisfacción, que guarda alguna semejanza con la noción de correspondencia. Compárese «X es verdadera si y sólo si existe un hecho S tal que X se corresponde con S» con «X es verdadera si y sólo si existe una secuencia S tal que X es satisfecha por S». Como la de correspondencia, la de satisfacción no es una relación simétrica y vincula oraciones con entidades extralingüísticas. La naturaleza de esas entidades, sin embargo, es problemática. Estrictamente son objetos matemáticos cuya existencia extramental puede ser puesta en duda y, en cualquier caso, no es de la misma especie que la que intuitivamente se esperaría de los objetos involucrados, en el caso de que las oraciones relacionadas con ellos fuesen empíricas. Tampoco se aclara la constitución de los objetos que integran las secuencias, y nada obsta para que se los considere de índole mental. En favor de las secuencias parece hablar el que, desde siempre, los hechos que ellas vienen a reemplazar fueron metafísicamente sospechosos. Lo cual sugiere la conveniencia de no acentuar su papel en la caracterización tradicional de la correspondencia, cediendo su lugar al concepto de *alguna* entidad extralingüística, y mejorando entonces las perspectivas de la reconstrucción tarskiana. Pero sí es claro, por lo antedicho,

que la definición es ontológicamente neutral en la disputa realismo-antirrealismo. Debemos distinguir esta observación de la que el mismo Tarski efectuó sobre la neutralidad gnoseológica de su definición (Tarski, 1944, § 18), señalamiento que no es sino una reiteración de que su tarea era la de analizar el concepto de verdad, y no la de proporcionar algún criterio epistémicamente útil para su aplicación.

Recientemente, se ha preferido remitir la cuestión de si un discurso es o no es realista, al esclarecimiento de la validez o invalidez del principio de bivalencia para las oraciones en cuestión. Según este punto de vista, las definiciones tarskianas comprometen con el realismo, ya que permiten deducir la bivalencia. Sin embargo, definiciones tarskianas son las que se obtienen mediante cierto método aplicado a un lenguaje *interpretado*. Entre las expresiones interpretadas, desde luego, estarán los signos lógicos, cuya interpretación se revela en las cláusulas específicas de la definición de satisfacción. Cláusulas que son, precisamente, las responsables de la derivación de la bivalencia en los ejemplos usados por Tarski. De modo que la bivalencia, en estos casos, no es consecuencia del método tarskiano sino de la índole del lenguaje al que se aplicó. Si la interpretación correcta de los signos lógicos de cierto lenguaje fuese, digamos, intuicionista, entonces la definición tarskiana de su predicado veritativo no implicaría bivalencia (véase Putnam, 1978, 2.ª conferencia, y las construcciones tarskianas para lenguajes con vacíos de valor veritativo, como en Kripke, 1975).

La relación tradicional de correspondencia descansa en una isomorfía entre oraciones y entidades externas o, por lo menos, en una asociación global entre esos objetos. Pero es también un nexo de fundamentación, desde aspectos del mundo (o desde el mundo como totalidad) hacia las oraciones y no, en general, a la inversa (a esto remite el criterio (F) del § III). La concepción defendida por Tarski muestra algo de ese rasgo de nuestro concepto corriente resumido en la tesis de la determinación de la verdad por parte del mundo. Lo hace cuando implica que *si* 'x gira' es satisfecha por los objetos que giran y 'Marte' es satisfecha por Marte, *entonces* si 'Marte gira' es satisfecha por todo objeto (esto es, si Marte gira) entonces 'Marte gira' es verdadera. Pero no dice todo lo necesario para establecer la relación de *determinación* desde el mundo a la oración. Porque esos condicionales son muy débiles y su definición de verdad permite agregar: *si* 'x gira' es satisfecha por los objetos que giran y 'Marte' por Marte, *entonces* si 'Marte gira' es verdadera, entonces Marte gira. Es justamente este rasgo de la *adaequatio* el que daría alguna justificación para restringir el criterio T de modo de evitar las paradojas del tipo

de la del mentiroso. El valor veritativo de una oración sólo podría establecerse de modo unívoco cuando hubiere algo —en general distinto de ella misma— que fundamentase la atribución (véanse Kripke y, antes, Herzberger). Para expresiones como 'Esta oración es falsa' o 'Esta oración es verdadera' la búsqueda de fundamento es infructuosa: al no poder eliminar la presencia de los predicados veritativos, no se alcanza a describir ningún hecho, independiente de ellas, que legitime alguna atribución de valor veritativo. La suma de autorreferencia y predicación de falsedad (o de verdad) es insostenible en estos casos. Es indiferente aquí que la primera expresión conduzca a contradicción pero la segunda no lo haga; y que tal vez pueda encontrarse alguna razón para aceptar 'Esta oración es verdadera', pero no pueda haberla para 'Esta oración es falsa'.

No siempre lo que verifica es independiente de la oración. Ser una oración no es un rasgo externo a las oraciones, tómese 'Ésta es una oración' y se tendrá un ejemplo de oración cuya verdad se funda en la propia oración. Pero no es éste un caso donde la verdad de una oración sea su propio fundamento. En ese ejemplo, la predicación de 'es una oración' generó a la vez la oración y aquello que la verifica. 'Ésta es una oración' es verdadera si y sólo si 'Esta es una oración' es una oración. Esa oración, pues, no se «trasciende» para ser verdadera o falsa. Pero 'es una oración' no es un término que «apunte» más allá de las oraciones, a diferencia de 'es verdadera', cuya predicación genera 'Esta oración es verdadera' pero no genera aquello que pueda verificarla. Si quisiéramos determinar sus condiciones veritativas sólo lograríamos establecer que 'Esta oración es verdadera' es verdadera si y sólo si 'Esta oración es verdadera' es verdadera. Con lo cual regresamos al punto de partida. Tenemos así algún motivo, enraizado en la idea (ahora correspondentista) de verdad, para comenzar a justificar el confinamiento del predicado veritativo al metalenguaje. Esta limitación es un modo —quizá brutal pero efectivo— de impedir la generación de oraciones incapaces de «trascenderse» en dirección al mundo externo a ellas mismas al que aparentan señalar.

Pero la limitación —alentada por Tarski— a un metalenguaje extensional (al menos en su aparato de análisis semántico de lenguajes-objeto) impide reconstruir cabalmente el concepto, intuitivamente intensional, de fundamentación. La mejor aproximación elemental habrá de conservar, al menos, su carácter no simétrico. Tal logra la noción de satisfacción, candidata obvia a suceder, en la teoría, a la correspondencia preteórica. La mera asimetría, es claro, no otorga preeminencia ontológica a ninguno de los términos intervinientes (en particular, no al primero en el orden, como lo muestra el que también

la relación conversa podría usarse con éxito); una teoría tarskiana y correspondentista deberá establecerlo de modo explícito.

Suprimir la mención de hechos puede contentar a quien sospeche que no existen o no encuentre un modo prístino de caracterizarlos (en particular, uno que no dependa de la predicación de verdad), esto es, puede verse como una mejora de la posición tradicional. Las sucesiones de objetos que los suplantan no aspiran a ninguna isomorfía con las oraciones; ni siquiera representan aspectos parciales del mundo que tengan el rol de fundar, específicamente, el valor veritativo de cada oración. Están, por así decir, representando al mundo en su totalidad. Cada sucesión selecciona objetos del mundo, pero *todas* satisfacen por igual a las oraciones verdaderas (y todas dejan de satisfacer, democráticamente, a las falsas).

Cambiando por un existencial el cuantificador de la definición tarskiana se lograría, sin perder ningún resultado deductivo, como ya vimos, cierta impresión de especificidad. Sobre todo si, en cada caso, de algún modo se privilegiaran aquellas sucesiones que sólo tienen como miembros a los objetos intuitivamente referidos por las expresiones componentes de las oraciones (repitiendo eternamente alguno, a fin de hacerlas infinitas). La artificiosidad del procedimiento, sin embargo, es inocultable. La «teoría», sin agregados explícitos (*i.e.*, sin dejar de ser un mecanismo técnico para generar los casos de T), no alcanza para distinguir ninguna secuencia particular. La secuencia cuyo primer lugar lo ocupa el tercer lunes de noviembre de 1993 en Corrientes, y el resto está ocupado por el Aconcagua (llamémosla δ), satisface tanto la oración 'La Luna gira', cuanto 'Colón llegó a Castilla' y, desde luego, cualquier otra oración verdadera. Se sigue que todas las secuencias satisfacen esas oraciones. De modo que la definición general de verdad podría haber sido

X es verdadera sss δ satisface X

y lo mismo valdrá con cualquier elección particular. La estructura de una oración no se refleja especialmente en la entidad (la totalidad de las secuencias o del mundo) que la satisface a ella y a todas las que se satisfacen. Su especificidad sólo interviene para la determinación de si se corresponde o no con esa entidad. Esta determinación, por su parte, no implica el reconocimiento de hechos (no implica mencionar ni cuantificar sobre hechos), ni siquiera el de propiedades o relaciones salvo, claro, que el lenguaje-objeto mismo lo haga, pero esto no es imputable a la noción de verdad así definida, sino sólo a la interpretación del lenguaje-objeto (si 'satisface' se definiese explícitamente,

como deseaba Tarski, la definición de verdad involucraría compromiso con la existencia de clases; pero podemos ser menos ambiciosos). Quien sólo esté interesado en la extensión del predicado veritativo, en saber si una oración dada es verdadera o, especialmente, en saber cómo se relaciona su valor veritativo con el valor veritativo de las demás, agradecerá la sencillez de la definición tarskiana, que determina las condiciones para la verdad y falsedad de las oraciones. Un ejemplo lo ofrecen aquellos interesados en efectuar análisis lógicos de los discursos. Quien pretenda saber, además, en qué consiste la verdad de cada oración, lamentará que la estructura de las consideraciones específicas que tienen que ser hechas a los efectos de aquella determinación no formen parte de la definición del concepto de verdad. Lamentará que la definición del concepto sólo permita exhibir sin decir los rasgos estructurales que determinan el valor veritativo.

Otra incomodidad, relacionada con la anterior, la suscita el empleo de secuencias infinitas en la definición de 'satisface'. Tarski, alertado, esbozó un recurso alternativo (Tarski, 1956, 195, nota 1). La incomodidad surge por la presencia de términos intuitivamente superfluos para la satisfacción de una cuasi-oración que, en todos los casos, sólo emplea un número finito de variables. La alternativa consiste en reemplazarlas por secuencias finitas, integradas solamente por aquellos miembros de cada secuencia infinita que ocupan los lugares señalados por los índices de las variables que aparezcan en la cuasi-oración (si ésta tenía n lugares, la secuencia tendrá, a lo sumo, n términos). El precio es un incremento en la complejidad técnica de la definición de satisfación, que estaría compensado por el mayor apego a la concepción intuitiva. Pero esto es ilusorio. Dado que las oraciones son casos extremos de cuasi-oraciones (aquellos en que el número de variables libres es cero), la definición de verdad debería ser:

X es verdadera sss la secuencia nula satisface X

Que sea el mundo globalmente, y no alguno de sus presuntos hechos parciales, lo que aparezca satisfaciendo una oración verdadera, ya era conceptualmente inquietante para el correspondentista. Que el conjunto vacío lo sustituya agrava intolerablemente su malestar (aunque ¿por qué el ser y no la nada? después de todo). Una persona tradicional, pero precavida ante los hechos, solicitará complementar la construción tarskiana con algún argumento que justifique distinguir, de entre todas las secuencias satisfactoras, aquella (o aquellas) que haga justicia a la creencia preteórica (a veces llamada «intuición») de que no todos los objetos intervienen en la constitución de cada obje-

to y que, entonces, algunos podrían haber existido en ausencia de algunos otros y, por ello, no todas las secuencias tarskianamente satisfactoras representan igualmente aquello en lo real que fundamenta la verdad de una oración particular. Pero un tradicionalista férreo observará que, en lugar de buscar motivos *ad hoc* para preferir alguna de estas definiciones alternativas, es mejor retornar a la tesis de la existencia de hechos para el papel de verificadores de oraciones (aunque no necesariamente para oficiar de entidades nombradas por ellas, como podría requerir alguna poco recomendable explicación del funcionamiento del lenguaje). Y supondrá que las entidades matemáticas aducidas en las definiciones (secuencias de algún tipo) sólo son sus representantes para simplificar la técnica de obtención de resultados (asignación de valores veritativos, demostración de principios generales como bivalencia, trivalencia, etc.). Al implicar los casos de T, la definición tarskiana señala condiciones veritativas específicas para oraciones específicas. ¿Qué diferencia a las condiciones veritativas de P, del denostado hecho representado por P? Nada interesante, *si* se acepta que las oraciones representan o que *hay* algo específico en el mundo que las hace verdaderas. Tarski muestra que podemos explicar el uso del predicado veritativo sin necesidad de atribuir a los usuarios la creencia de que existen hechos (aunque sí es necesario atribuirles la creencia de que existen objetos). Pero el que trata de aclarar la noción de verdad no está, por eso, intentando explicar, económicamente, la conducta lingüística.

Observaciones como las anteriores ofrecen un indicio más de que, al construir su definición, Tarski confió indebidamente en la fuerza del criterio T respecto del objetivo de reconstruir la idea correspondentista tradicional. Sus secuencias no son hechos, su relación de satisfacción no establece un nexo suficientemente fuerte entre oraciones y mundo (y sólo por su asimetría simula el vínculo de fundamentación), y su mundo podría ser una mera construcción mental. Pese a todo, sus definiciones lo pusieron más cerca de la correspondencia aristotélica que donde el esquema T lo había dejado. Obsérvese que si bien coincide con la teoría mínima al garantizar el criterio T, el método tarskiano enfrenta, en lugar de eludir, la cuestión de las antinomias, y que las definiciones tarskianas, a diferencia de las teorías mínimas, otorgan un papel teórico al valor semántico de los componentes suboracionales, a través del concepto de satisfacción. Esta diferencia tiene interesantes consecuencias, ya que las teorías mínimas no pueden justificar afirmaciones del tipo: «cualesquiera sean las oraciones P y Q, 'P & Q' es verdadera exactamente cuando 'P' es verdadera y 'Q' es verdadera». Las teorías mínimas implican cada una de las

instancias de ese esquema, pero, para la generalización, requiere la postulación adicional de que las que implica son todas.

Las definiciones tarskianas, en suma, *no son realistas, no son correspondentistas al modo clásico* (no necesitan apelar a hechos) y, en la medida en que no legitiman el criterio (F), *no son un análisis adecuado* del concepto «tradicional» de verdad.

7. REFERENCIA Y VERDAD

Es natural pensar que en el examen del valor semántico que las expresiones suboracionales tengan, se encontrará una vía para sostener la conjetura de que el valor veritativo de una oración aislada depende especialmente de algunos aspectos de la realidad, y no de todos o, al menos, no de todos por igual. Que P sea verdadera de este mundo, depende de cuál sea éste; lo que a su vez depende, por así decir, de todos los aspectos del mundo. Pero mucho de lo que hacemos y pensamos supone la creencia de que, en lo esencial, ciertos aspectos del mundo (entidades, relaciones, circunstancias, sucesos) son independientes de otros, por lo menos en tanto nos afectan. La intención de voto de mi prima, el asteroide más pequeño de la constelación de Orión y la ardilla más longeva de las nacidas en 1689 caracterizan este mundo tanto como sus respectivos ser penosa, no haber sido descubierto hasta ahora y haber muerto en el siglo 18. Pero, en lo que nos importa, este mundo real podría ser uno en el que algo de esto no hubiese ocurido. Diderot y el grado de su erudición, por ejemplo, seguirían siendo los mismos, solemos creer, tanto si la ardilla murió en su siglo o en el anterior, o si mi prima no tuviera las preferencias que tiene. En el nivel de detalle en que conocemos el mundo, es enteramente plausible la hipótesis de que esos aspectos son independientes, esto es, podrían pertencer a diferentes mundos y también, *a fortiori*, que nada decisivo para su caracterización individual los une en los mundos donde estén juntos.

Según la vieja idea correspondentista, la verdad de 'Colón llegó a Castilla' es independiente de mucho de lo ocurrido en el mundo, pero depende de las entidades de *este* mundo que estén relacionadas con los términos 'Colón', 'llegó a' y 'Castilla'. Quizás ignoremos cuáles sean exactamente pero, los correspondentistas creen, de ellas deriva el valor veritativo de oraciones como la citada; y podemos ir mejorando nuestra identificación de sus rasgos esenciales, al menos podemos distinguirlas de muchas otras (como ardillas longevas o partidos políticos argentinos) cuya contribución al valor veritativo de esa oración es despreciable.

El valor semántico de los componentes suboracionales de una oración X debe tomarse en cuenta para la determinación tarskiana de las condiciones veritativas de X (esto ocurre en primer lugar, y de modo implícito, al establecer la noción de 'σ satisface φ' para las oraciones o las funciones proposicionales carentes de signos lógicos). Es en virtud de que 'I' está asociado con la relación de inclusión entre conjuntos y 'v_k' y 'v_l' están asociados, relativamente a σ, a los conjuntos A y B, que 'Iv_kv_l' es satisfecha por σ si y sólo si A está incluido en B. Si el lenguaje contuviese constantes, digamos 'a', 'b', ..., 'n', que hiciesen referencia, respectivamente, a los conjuntos A, B, ..., N, entonces sería debido a estas asociaciones que 'Iab' es satisfecha por σ si y sólo si A está incluido en B. Pero estas especificidades se ocultan al pasar a la definición de verdad. '(Ev_kv_l) Iv_kv_l' y 'Iab' son verdaderas si y sólo si (y sólo ahora estamos frente a una equivalencia definicional) toda secuencia las satisface, que es tanto como decir que ningún ordenamiento de los objetos del mundo suministra un par ordenado de objetos correspondientes a 'v_k' y 'v_l', o a 'a' y 'b' que impida el cumplimiento de las condiciones *de satisfacción* de esas oraciones. Pero es menos que decir que su verdad *consiste* en la existencia de dos conjuntos uno de los cuales está incluido en el otro o, respectivamente, que consiste en que el conjunto referido por 'a' está incluido en el conjunto referido por 'b'. Y está lejos de afirmar, de modo más general, que su valor veritativo depende exactamente de que los objetos referidos por los términos singulares mantengan la relación asociada con el término general.

Pero aquí conviene una aclaración en favor del carácter elucidatorio (incompleto) de las definiciones tarskianas. La contribución principal del trabajo de Tarski reside en: 1) su demostración, por el ejemplo, de la posibilidad de definir adecuadamente el predicado veritativo característico de un lenguaje de suficiente poder expresivo y, 2) la sugerencia de que el camino seguido en su ejemplo da sustento a la idea de un método general para construir esa clase de definiciones. «La descripción abstracta general de este método y de los lenguajes a los que fuese aplicable sería problemática y no del todo clara» (Tarski, 1956, 168). El lenguaje-objeto es, como se dijo, interpretado; y el método aludido incluye el conocimiento de esta interpretación. Debemos saber que 'I' expresa la relación de inclusión y que las variables remiten a conjuntos. En general, debemos ser capaces de reconocer oraciones, pero también términos singulares, generales, predicados, functores, signos lógicos, e identificar sus correspondientes referencias o extensiones o valores semánticos específicos. Sobre esta base se ha de construir una definición de 'σ satisface φ' del tipo de la

ejemplificada por Tarski y, finalmente, puede ofrecerse una definición tarskiana de 'es verdadera'. Muchas veces se ha objetado el valor filosófico de una elucidación de esta clase, observando que el predicado definido es, estrictamente, 'verdadera-en-L', donde 'L' es un «accidente sintáctico» tanto como 'verdadera'. La relativización a un lenguaje es, en un sentido, obvia, ya que el predicado veritativo lo es de oraciones y éstas son siempre relativas a un lenguaje, de modo que lo definido no es tanto 'X es verdadera' (o 'X es verdadera en L') cuanto la frase 'X de L es verdadera'. Cuando 'L' pueda cuantificarse se podría buscar una definición de la forma '(X)(L) (X de L es verdadera si y sólo si ...)' donde 'es verdadera' no quedaría relativizado a ningún lenguaje particular. De modo que, se dice, sólo para 'L' variable cabría admitir que lo definido es el concepto preteórico de verdad, y sólo entonces se contaría con una genuina aclaración. Sin embargo, el método definicional bosquejado antes es el sucedáneo tarskiano de este concepto general, concepto cuya definición consistente, hemos visto, es imposible mediante el empleo estricto de ese mismo procedimiento. Tarski prefirió no elaborar formalmente la idea del método, sino que intentó sugerirla a través de un ejemplo sencillo. Si recordamos que una elucidación de interés filosófico no necesita exhibir la forma de una definición explícita (ver § III), tendremos motivo, a pesar de la objeción citada, para otorgarle méritos aclaratorios al bosquejo aludido. Y, eso, aun antes de pretender que al requerir la capacidad de interpretar cada X, queda eximida la remisión explícita a un lenguaje particular.

Observemos, por otra parte, que una definición tarskiana mínima para un lenguaje particular es suficiente a los fines de fundamentar las reglas inferenciales vigentes en cierta práctica lingüística (la concerniente a ese lenguaje) o para evaluar sistemas lógicos alternativos. Una definición de esta clase basta para mostrar las relaciones de valor veritativo entre oraciones y cuasi-oraciones simples y complejas (compuestas y cuantificadas) que dependan de nexos lógico-formales del tipo de los representados por los signos lógicos elementales. Para estos fines no importa la materia especial de las oraciones, sólo su «forma lógica» y el que posean algún valor veritativo, independientemente de las razones que expliquen el valor veritativo que de hecho posean. No obstante, si el objetivo prioritario del análisis no es la lógica subyacente a una práctica lingüística, sino la idea de verdad, una definición de aquella especie resultará insuficiente (al menos para el correspondentista tradicional). Y también conviene recordar que no es el concepto mínimo de verdad, sino algún otro, el que interviene en la justificación usual de la aplicabilidad de los sistemas de lógica al lenguaje

natural (a través de la intención representacional atribuida al discurso cognoscitivo normal).

Cuando en la elucidación de la noción de verdad se hacen intervenir los conceptos de referencia de los términos singulares y de extensión de los predicados, vale decir, la creencia de que existen entidades referidas por los términos singulares y grupos de entidades representados por los predicados, pueden explícitamente distinguirse diferentes aspectos del mundo como verificadores de, digamos, 'Colón llegó a Castilla' y 'La luna gira' (aquí, apelando a ambos conceptos), y también pueden especificarse diferentes verificadores para 'Colón llegó a Castilla' y 'Colón se fue de Castilla' (recurriendo ahora a la idea de extensión de un predicado). Hacer depender de modo explícito la noción de verdad de la noción de referencia (tal como se defiende en Field, 1972) permite generalizaciones del tipo: una oración de un lenguaje L constituida por un predicado monádico aplicado a un nombre propio es verdadera si y sólo si el objeto referido por el nombre propio tiene la propiedad (o pertenece a la clase) referida por el predicado. Estos enunciados satisfacen las «intuiciones» correspondentistas clásicas más acabadamente que las afirmaciones como 'Fa' de L es verdadera si y sólo si toda secuencia la satisface, o 'Fa' de L es verdadera si y sólo si no existe una secuencia tal que A no es φ (que reclaman aclarar que 'φ' y 'A' traducen 'F' y 'a', o que a es A y los F son los φ; y en las que la noción de secuencia de objetos parece completamente irrelevante). En buena medida estas últimas sirven para entender (o exponer en qué consiste entender) la oración 'Fa'; pero no bastan para entender 'Fa es verdadera', si la verdad se entiende al modo correspondentista como «genuina» propiedad de oraciones. Precisamente esta capacidad de dar sentido (extensional) que exhiben esos bicondicionales fue rápidamente empleada en la exposición y fundamentación de sistemas lógicos construidos sobre la base de «lenguajes» caracterizados por reglas puramente sintácticas. Tarski dio por sentada la comprensión del lenguaje a fin de entender la idea de verdad, pero sus construcciones formales fueron usadas, canónicamente (véanse los textos de lógica de uso corriente), para dotar de significado a sistemas puramente sintácticos, a efectos de asemejarlos apropiadamente (para fines lógicos) a los lenguajes naturales. De esta semilla, seguramente, creció la semántica davidsoniana para los lenguajes naturales: al poner entre paréntesis la intelección de un lenguaje cualquiera, buena parte de lo que queda es un sistema de conjeturas sintácticas, y los procedimientos debidos a Tarski son una interesante promesa de restitución de lo que falta.

VIII. PALABRAS, ORACIONES, REALIDADES

Estrechar nexos conceptuales entre verdad y referencia conviene al correspondentista, pero también lo compromete a reflexionar sobre la idea misma de referencia. Las asociaciones entre palabras y objetos pueden entenderse de varios modos; según uno de ellos, son meros instrumentos metodológicos para asignar condiciones veritativas a las oraciones (es decir, para «demostrar» los bicondicionales T), objetivo alcanzable con un número indefinido de esquemas referenciales distintos (que no harían sino mostrar la irrealidad esencial de cada uno de ellos). Según el modo que elegirá el correspondentista, en cambio, serán conjeturas empíricas acerca de un fenómeno real, el de la efectiva conexión entre usos de expresiones suboracionales y aspectos y circunstancias del mundo. Circunstancias que tal vez no sean individualizables con exactitud (o con independencia de los hablantes), lo que daría lugar a diferentes esquemas referenciales hipotéticos, pero que de hecho existen y son «responsables» de que los hablantes hayan asociado con ellas sus palabras. A los fines prácticos, en ciertos contextos, 'gavagai' puede referirse indiferentemente a conejos o patas de conejo no separadas (aunque seguramente no se referirá a revoluciones políticas del siglo XIX), pero alguna entidad relacionada con esas descripciones —se dirá— está de hecho asociada con la expresión, en virtud de un fenómeno complejo desencadenado por la eficacia causal de algunas de sus propiedades.

Esta manera de entender la noción de referencia, en términos de los efectos que ciertas entidades causan en los partícipes de un lenguaje (defendida en Kripke, 1972, y Putnam, 1975) mejora la comprensión de esa idea. Pero, además, permite al correspondentista sustentar su creencia preteórica en la existencia de un nexo de fundamentación desde el mundo hacia el valor veritativo de las oraciones. En efecto, las propiedades habitualmente atribuidas a la relación causal son suficientes al respecto. Y, por añadidura histórica, el recurrir a la causalidad para este fin parece acordar con el modo en que el propio Aristóteles veía el problema (véase *Categorías,* 14b19-24 y *Metafísica,* Θ 1051).

Desde hace algún tiempo tienen éxito, contra la idea de referencia que resulta más afín al correspondentista, variados argumentos basados en la posibilidad de permutar las presuntas referencias de las palabras. Se inspiran en el procedimiento rutinario que los lógicos siguen para generar modelos distintos para una misma sintaxis. Por ejemplo, el «lenguaje de fondo» al que remite Quine para relativizar cualquier asignación de referencias puede utilizarse plenamente para hablar de sí mismo, pero entendido ahora como un sistema de meras

disposiciones a la conducta verbal. Disposiciones que, como la estructura puramente sintáctica de los lenguajes artificialmente construidos, reclaman del intérprete que les imagine conexiones con entidades extralingüísticas, permitiéndole, así, elaborar distintos esquemas referenciales compatibles con esas disposiciones e incompatibles entre sí. Una vía por la cual el correspondentista más clásico puede intentar sortear el problema comienza distinguiendo entre el objeto que el hablante cree estar nombrando (o los varios objetos que un observador puede creer que el hablante cree nombrar) y aquello del mundo que explica parcialmente el comportamiento lingüístico del hablante. El primer objeto interviene en la constitución de la significatividad de lo que el hablante dice y, cayendo bajo los argumentos de permutación, se revela equívoco y, eventualmente, innecesario. Pero el correspondentista postulará el segundo para determinar objetivamente las condiciones de verdad de lo que el hablante dice, lo propondrá como fundamento no epistémico del valor veritativo que de hecho tienen las oraciones que emita, no para caracterizar la significatividad de sus emisiones (problema que no es el suyo y para el cual, podrá conceder, no se requiere un objeto específico). Algunas personas han pensado que este enfoque fracasa porque las mismas dudas y equivocidades que suscita la referencia de 'relación referencial' las presenta la de 'relación causal' con que se las quiere suprimir. Claro que 'las mismas dudas' suscita las mismas dudas, al igual que lo demás que digamos. Esta crítica, pues, corre el riesgo de sumirnos en el silencio sin sentido. Entonces, según aquella perspectiva correspondentista (más problemática de lo que aquí parece), comprender ciertas prácticas lingüísticas (en el sentido de poder participar de ellas) no alcanza *per se* para determinar «la» estructura de la realidad. Pero que lo real no tiene por qué venir definido como subproducto de la explicación más restringida que demos de nuestro comportamiento semántico ya era, después de todo, una idea básica del correspondentista tradicional.

Se ha sugerido una manera diferente de expresar la dependencia de la atribución de verdad respecto de la realidad, empleando algo del enfoque semántico desarrollado por los lógicos para tratar los enunciados modales (Alchourrón, 1993). Según esto, lo que debe definirse es el predicado 'X es verdadera respecto de W', pudiendo entenderse que 'W' varía sobre el conjunto de los estados globales de la realidad en los diversos instantes, o sobre el conjunto de los posibles estados globales de la realidad, permitiendo representar la relativización de la verdad oracional a los cambios reales o posibles. La condición de verdad de una oración simple, digamos 'Fa', resultaría: 'Fa' es verdadera en W si y sólo si el objeto de W referido por 'a' es de la

clase de objetos de W asociados a 'F'. Condición ésta derivada de la definición de verdad (ahora relativa a una «realidad particular»): (X) (X es verdadera en W si y sólo si toda secuencia de objetos de W satisface X).

Sin embargo, cuando se entienden las oraciones como expresiones interpretadas, y la interpretación en términos de conexiones entre lenguaje y realidad (más que en términos de participación en el intercambio lingüístico), la sola consideración de una oración X es, *ipso facto*, la consideración de una «realidad particular»: la determinada parcialmente por las entidades directamente vinculadas con ciertas expresiones suboracionales de X. Por ejemplo, si X fuese 'Fa' con 'a' y 'F' asociados al objeto A y al conjunto de objetos φ, respectivamente, entonces, al margen de lo que crean los emisores de X acerca de φ, el análisis tarskiano muestra que si, y sólo si, A es uno de los objetos de φ, X tendrá la propiedad de ser verdadera. Debido a la influencia ejercida por el empleo del aparato tarskiano a fin de dotar de significado a un sistema sintáctico (o a fin de aclarar los significados en un lenguaje natural), puede sobresaltar el aparente hecho de la inevitabilidad de la verdad o falsedad de las oraciones. En efecto, si se ha establecido, *i.e.* se sabe verdadero, que A es el objeto referido por 'a' y todos los objetos agrupados por 'F' se conocen como tales, ninguna verdad ni falsedad parece contingente. Y surgirá la idea de hacer explícita referencia a realidades alternativas (mundos posibles) en el concepto de verdad. Esto es, la idea de que el concepto preteórico a elucidar no es 'X es verdadera' sino 'X es verdadera en W' (donde la relativización a mundos diversos se exhibe ya en el *definiendum*).

Pero cuando, soslayando aquella influencia, la construcción tarskiana se usa para aclarar la idea de verdad (no la de significado), nada obsta para que ese empleo presuponga la comprensión del lenguaje-objeto y para creer que esta comprensión no consiste en (ni incluye) el conocimiento exhaustivo de las referencias y las extensiones de todos los términos singulares y generales del lenguaje. Entonces no será problemático —siguiendo el concepto correspondentista-tarskiano— aceptar que, si *de hecho* A y φ son lo referido por 'a' y 'F' y si *de hecho* A es un φ, estarán dadas las condiciones necesarias y suficientes para la verdad de X, aun cuando las cosas podrían ser de otra manera. Si supiésemos que ésos son los hechos referenciales pertinentes, entonces nada más requeriríamos, en particular ninguna comprobación fáctica adicional, para saber que X es verdadera; pero ¿qué tiene esto de anómalo? Saber esos hechos equivale a saber la verdad de X, porque en eso consiste su verdad. Necesariamente, si a es A, los F son los φ y A es φ, entonces X es verdadera, pero esto no significa que

X sea necesariamente verdadera. Y, por otra parte, dado que el conocimiento del antecedente no es *a priori*, no se requiere acudir a la relativización a mundos posibles para restituir la contingencia ni la aposterioridad. La dependencia del mundo podrá venir expresada por el *definiens*, no hará falta cambiar el *definiendum* tradicional. El recurso a los mundos o estados de cosas posibles, sin embargo, queda sugerido inmediatamente cuando se trata de explicar la verdad de las oraciones explícitamente modales; pero esto depende de la intelección de los giros modales, no del concepto general de verdad. Quizás se estime, por otra parte, que una explicación completa del significado de 'verdadera' solicita aclarar detalladamente la índole de las asociaciones referenciales involucradas. El punto es discutible, pero, en todo caso, no se requiere esa aclaración para reflejar el prejuicio correspondentista tradicional de que la aplicación del predicado veritativo depende del mundo efectivo (real), pero éste, sea cual fuere, no tenía su existencia metafísicamente asegurada.

Una versión de la teoría de la correspondencia, usufructuaria de la técnica tarskiana, como la aludida en esta sección, tiene dos rasgos interesantes. Por una parte, permite liberar a la teoría de la presencia de entidades problemáticas como hechos negativos o condicionales o disyuntivos, para explicar las condiciones veritativas de negaciones, condicionales o disyunciones. Por otra parte, permite acoger la idea correspondentista más general —la de explicar la verdad en términos de relaciones entre lenguaje y realidad extralingüística— sin introducir entidades especiales correspondientes a las oraciones. Hace técnicamente prescindible la noción de hecho aun para las oraciones simples. Pero, es claro, no impide su empleo ni, menos aún, su representación mediante entidades abstractas. Para determinar las condiciones veritativas de una oración simple, no se requiere destacar otros objetos cuantificables más que los individuales. No hay necesidad de suponer que los predicados ocupan lugares accesibles a cuantificación ni, menos aún, imaginar entidades en lugar de las cuales estarían las oraciones mismas. Reconoceremos tres tipos de entidades lingüísticas categoremáticas: términos singulares, predicados y oraciones. Pero sólo hace falta considerar que las primeras refieren a entidades extralingüísticas; sobre las segundas basta creer que se aplican a (o son verdaderas de) algunas de esas entidades, y de las terceras, a lo sumo, cabrá decir que todas o ninguna de esas entidades las satisfacen.

Quienes crean que las entidades básicas son los hechos y que los objetos singulares, las propiedades y los conjuntos no son sino modos epistémicamente útiles para comparar y distinguir hechos, o quienes simplemente crean que hay hechos además de otras cosas, no tendrán

interés filosófico en aprovechar esa facilidad técnica. Pero en ciertos círculos son más los que prefieren pensar teorías complejas sobre entidades simples y pocas que elaborar, sencillamente, teorías sobre el mundo. Y entonces siguen de cerca la construcción tarskiana y desdeñan los hechos. A veces se esgrimen, no obstante, motivos menos temperamentales. Por ejemplo, que los hechos no pueden identificarse sino como las entidades asociadas a oraciones verdaderas, con lo cual se produce un círculo conceptual sombrío. O, más dramáticamente, que existe a lo sumo un hecho. La primera objeción se emparienta con otra, dirigida contra la idea de correspondencia, basada en la imposibilidad de comparar oraciones con algo distinto de oraciones. Esta crítica —cabe responder— confunde el método para creer que una oración es verdadera con el concepto de su verdad (véase § II). De manera semejante puede replicarse a aquélla, diferenciando entre un sentido epistémico de «identificar», que alude a los recursos con que contamos para reconocer y distinguir unas entidades de otras, y un sentido definicional, que remite al criterio para distinguirlas conceptualmente. Por otra parte, la tesis de la imposibilidad de la existencia de más de un hecho, que se hace remontar dificultosamente hasta Frege, se afirma sobre la base de premisas de sospechosa simplicidad (véase Barwise y Perry, 1981). En este contexto importa señalar, simplemente, que la vaciedad del concepto de hecho es tan poco obvia como su riqueza.

El núcleo de las teorías correspondentistas tradicionales, expresado con la debida imprecisión, reside en creer que la verdad de una oración depende (parcialmente) de un factor irreductible a las capacidades de quien la emite (sean sus criterios de aceptabilidad, o de satisfacción de necesidades, o cualquier circunstancia característicamente subjetiva). Este factor «objetivo» juega un papel en la explicación de la revisabilidad de las creencias. Qué sea esta objetividad, es problemático. Clásicamente solía resguardarse en los hechos en sí mismos o en el mundo en sí. Postkantianamente parece recluida en los fenómenos intersubjetivamente (a veces lingüísticamente) constituidos (a veces construidos). Sin duda, versiones como las últimas cuestionan la independencia (aunque tal vez no la irreductibilidad) de ese factor objetivo y, con eso, moderan los ímpetus del realismo espontáneo (lo envejecen). Y ya se observó (§ VI) que la variante tarskiana de la correspondencia ni siquiera excluye, *per se*, una posición francamente no realista. Pero este problema no es, *prima facie*, el del concepto de verdad (al menos no el del funcionamiento de esa noción en la trama conceptual preteóricamente compartida) sino el de la noción misma de realidad. Reflexionar sobre uno puede ser inseparable de examinar

el otro, pero, en cualquier caso, será esclarecedor conocer cuánto de la idea de verdad, de su papel en el intercambio lingüístico y en la estructuración del pensamiento, cabe conjeturar antes de avanzar (si acaso) en la elaboración del concepto de lo real*.

REFERENCIAS BIBLIOGRÁFICAS

Alchourrón, C. (1993), «Concepciones de la lógica», en Íd., *Enciclopedia Iberoamericana de Filosofía*, vol. 7: *Lógica*, Trotta/CSIC, Madrid, 1995.

Austin, J. L. (1950), «Truth»: *Proceedings of the Aristotelian Society*, sup. vol. 24.

Barwise, J. y Perry, J. (1981), «Semantic Innocence and Uncompromising Situations», en French, Uehling y Wettstein (comps.), *Midwest Studies in Philosophy*, vol. 6, Minneapolis, University of Minnesota Press.

Dummett, M. (1959), «Truth»: *Proceedings of the Aristotelian Society*, vol. LIX. (Trad. castellana en Dummett, *La verdad y otros enigmas*, FCE, México, 1990.)

Ezorsky, G. (1967), «Pragmatic Theory of Truth», en Edwards (comp.), *Encyclopedia of Philosophy.*

Field, H. (1972), «Tarski's theory of truth»: *Journal of Philosophy*, LXIX, 13.

Frege, G. (1918), «Der Gedanke», en sus *Kleine Schriften*, comp. por I. Angelelli, Darmstadt, 1967. (Trad. castellana de A. Gómez-Lobo, en *Siete escritos sobre lógica y semántica*, Valparaíso, 1972.)

Grover, D. (1992), *A Prosentential Theory of Truth,* Princeton University Press.

Hempel, C. (1935), «On the Logical Positivists' Theory of Truth»: *Analysis*, vol. 2, n.º 4.

Hesse, M. (1980), «Habermas' Consensus Theory of Truth», en su *Revolutions and Reconstruccions in the Philosophy of Science*, Bloomington, University Press.

Hochberg, H. (1992), «Moore's anticipation of Tarski's Convention T and his refutation of truth as coherence»: *History of Philosophy Quarterly*, vol. 9, n.º 1.

Horwich, P. (1990), *Truth*, Blackwell, Oxford.

Kripke, S. (1972), «Naming and necessity», en Davidson y Harman (comps.), *Semantics of Natural Language*, Reidel, Dordrecht. (Trad. castellana de M. Valdés [de la versión como libro de 1981], *El nombrar y la necesidad*, México, UNAM, 1985.)

Kripke, S. (1975), «Outline of a theory of truth»: *Journal of Philosophy*, 72. (Trad. castellana de M. Valdés en *Cuadernos de Crítica*, n.º 36, México, UNAM, 1984.)

* La primera versión de este texto se redactó en agosto de 1994 para formar parte de este volumen, y se imprimió, dejando la debida constancia, como publicación interna de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires, en febrero de 1996.

Putnam, H. (1975), «The meaning of meaning», en Gunderson (comp.), *Language, Mind and Knowledge*, Minneapolis, University of Minnesota Press. (Trad. castellana de J. Flematti en *Cuadernos de Crítica*, n.º 28, México, UNAM, 1984.)
Putnam, H. (1978), *Meaning and the Moral Sciences*, Routledge, London. (Trad. castellana de A. I. Stellino: *El significado y las ciencias morales*, México, UNAM, 1991.)
Quine, W. (1970), *Philosophy of Logic,* Harvard U. P., Cambridge/London. (Trad. castellana de M. Sacristán: *Filosofía de la lógica*, Madrid, Alianza, 1973.)
Ramsey, F. (1927), «Facts and propositions»: *Proceedings of the Aristotelian Society,* supp. vol. 7.
Russell, B. (1912), *The Problems of Philosophy,* Oxford University Press.
Tarski, A. (1936), «Der Wartheitsbegriff in den formalisierten Sprachen»: *Studia Philosophica*, vol. I (versión inglesa en Tarski, 1956).
Tarski, A. (1944), «The Semantic Conception of Truth»: *Philosophy and Phenomenologial Research,* vol. IV. (Trad. castellana de M. Bunge, en Bunge [comp.], *Antología semántica*, Buenos Aires, Nueva Visión, 1960.)
Tarski, A. (1956), *Logic, Semantics and Metamathematics*, Clarendon, Oxford.
Tarski, A. (1969), «Truth and Proof»: *Scientific American*, vol. 220, n.º 6.
White, A. (1957), «Truth as Appraisal»: *Mind* 66.

La bibliografía lógico-semántica sobre la noción de verdad se ha incrementado notoriamente en los últimos años. Son especialmente útiles:

Blackburn, S. y Simmons, K. (eds.), *Truth*, Oxford University Press, 1999.
García Carpintero, M. (1996), «What is a Tarskian Theory of Truth?»: *Philosophical Studies,* 82.
Nicolás, J. A. y Frápolli, M. J. (comps.) (1997), *Teorías de la verdad en el siglo* XX, Tecnos, Madrid.

LA NOCIÓN DE CONSECUENCIA LÓGICA

Mario Gómez Torrente

1. CONSECUENCIA LÓGICA, NECESIDAD LÓGICA, FORMA LÓGICA

Suele decirse que la lógica se ocupa de caracterizar la noción de argumento correcto, o sea, de argumento cuya conclusión es consecuencia de sus premisas (y de buscar formas de separar de manera práctica los argumentos correctos de los argumentos incorrectos). Pero eso es una manera condensada de expresar la verdad. La verdad es que los objetivos de la lógica, aun siendo ambiciosos, no lo son tanto. La lógica, al menos si entendemos este sustantivo de la manera más frecuente entre los filósofos, se ha ocupado tradicionalmente de caracterizar la noción de argumento *lógicamente* correcto, o sea, de argumento cuya conclusión es consecuencia *lógica* de sus premisas (y de buscar formas de separar de manera práctica los argumentos lógicamente correctos de los argumentos lógicamente incorrectos).

La noción intuitiva de consecuencia lógica posee un grado notable de vaguedad, y la línea que la acordona, por así decir, en el interior de la noción más general de consecuencia es una línea difusa. De todos modos, hay un par de rasgos intuitivos, también vagos, de la noción intuitiva de consecuencia lógica que a menudo permiten distinguir los argumentos lógicamente correctos de los meramente correctos. En primer lugar, si una conclusión se sigue lógicamente de unas premisas, entonces esa conclusión se sigue *por necesidad lógica* de esas premisas. No siempre que un filósofo dice que una conclusión se sigue (a secas) de unas premisas está dispuesto a decir también que esa conclusión se sigue por necesidad lógica de esas premisas. Por ejemplo, es quizá razonable aceptar que el argumento con

premisa 'Las leyes de la física no han tenido excepciones hasta hoy' y conclusión 'Mañana no habrá excepciones a las leyes de la física' es un argumento correcto. Pero no es un argumento lógicamente correcto, y una manera de convencerse de ello la tendrá quien tenga la intuición de que su conclusión no se sigue por necesidad lógica de sus premisas.

Aun si no tenemos una idea muy clara de qué es lo que hace que una conclusión se siga por necesidad lógica de unas premisas, es claro que cuando esto ocurre tiene que haber una conexión modal muy estricta entre las premisas y la conclusión (una conexión que no se da entre la premisa y la conclusión del ejemplo anterior). Seguramente la mayoría de los que han pensado sobre estos asuntos aceptarían que esa conexión modal se da, por ejemplo, entre una premisa y una conclusión con el mismo significado que esa premisa: 'Hay al menos uno que es cabuérnigo y es albéitar' se sigue por necesidad lógica de 'Hay al menos uno que es cabuérnigo y es veterinario'. Otros casos simples en que se da la conexión modal que tengo en mente son casos en los que en la premisa se califica afirmativamente a algunos con un adjetivo y en la conclusión se sustituye el adjetivo con un calificativo que explica parte del significado del adjetivo original: de 'Todo el que es sabio es soltero' se sigue por necesidad lógica 'Todo el que es sabio no es casado' (no estoy muy seguro de que 'es soltero' y 'no es casado' sean sinónimos tal como estas expresiones se usan normalmente, pero no me cabe duda de que todo soltero es necesariamente no casado).

Pero los argumentos del párrafo anterior, a pesar de ejemplificar seguramente la conexión modal que se da en los argumentos lógicamente correctos, no son en sí lógicamente correctos. Una manera de verlo es que no poseen el segundo rasgo intuitivo de la noción intuitiva de consecuencia lógica al que me he referido antes. Este segundo rasgo es el siguiente: si un argumento es lógicamente correcto, entonces *todo argumento con la misma forma lógica es también lógicamente correcto*. La noción de forma lógica es nuevamente muy vaga, pero algo se puede decir para intentar aclararla un poco. Lo primero que hay que decir, aunque sólo aclare un poco y no explique mucho, es que la forma lógica de un argumento es una cierta forma esquemática que tienen en común los argumentos que comparten forma lógica. Intuitivamente, el argumento con premisas 'Todo el que es sabio transige' y 'Todo el que transige es pesimista' y conclusión 'Todo el que es pesimista es sabio', que podemos representar de la manera habitual,

Todo el que es sabio transige
Todo el que transige es pesimista

Todo el que es pesimista es sabio,

comparte forma lógica con el argumento

Todo el que es felino es mamífero
Todo el que es mamífero es vertebrado

Todo el que es vertebrado es felino.

Hay una forma esquemática que comparten estos dos argumentos que resulta de sustituir de manera uniforme (con los mismos signos siempre sustituidos por los mismos signos) en uno cualquiera de ellos las *expresiones no lógicas* por letras esquemáticas sin significado propio. (No hago un intento de ser completamente preciso en la caracterización del método de sustitución una vez fijadas las expresiones no lógicas.) Esta forma esquemática es la forma lógica de esos argumentos. En el ejemplo la forma lógica que comparten los dos argumentos es la siguiente:

Todo el que *F G*
Todo el que *G H*

Todo el que *H F*,

que se obtiene, por ejemplo, sustituyendo en el primer argumento 'es sabio' por '*F*', 'transige' por '*G*' y 'es pesimista' por '*H*'. Es bueno notar que los argumentos de este ejemplo son lógicamente incorrectos, y una manera de convencerse de ello es convencerse de que las conclusiones no se siguen por necesidad lógica de los conjuntos de premisas correspondientes: por ejemplo, aun cuando todos los felinos son de hecho mamíferos y todos los mamíferos son de hecho vertebrados no se sigue por necesidad (lógica o no), y de hecho es falso, que no haya de haber vertebrados no felinos.

De nuevo intuitivamente, el primer argumento de hace dos párrafos, es decir,

Hay al menos uno que es cabuérnigo y es veterinario

Hay al menos uno que es cabuérnigo y es albéitar,

no es lógicamente correcto, aunque la conclusión se siga por necesidad lógica de la premisa. Una manera de convencerse de ello es convencerse de que no posee el segundo rasgo intuitivo de los argumentos lógicamente correctos: hay argumentos lógicamente incorrectos con la misma forma lógica. La forma lógica del argumento es

Hay al menos uno que *F* y *G*

Hay al menos uno que *F* y *H*.

Un argumento con esta misma forma lógica es el siguiente:

Hay al menos uno que es viudo y feliz

Hay al menos uno que es viudo y lactante.

Y este argumento es claramente lógicamente incorrecto, al menos por el hecho de que su conclusión no se sigue por necesidad lógica de su premisa: el que haya al menos un viudo feliz (y seguramente los hay) no implica por necesidad (lógica o no) que alguno de los viudos esté lactando (esto es con toda probabilidad falso, por muchas viudas lactantes que haya). Igualmente, el argumento

Todo el que es sabio es soltero

Todo el que es sabio no es casado

es lógicamente incorrecto, y una manera de convencerse de ello es convencerse de que tiene la forma lógica

Todo el que *F* *G*

Todo el que *F* no *H*

y de que hay argumentos que comparten esa forma lógica y en los cuales la conclusión no se sigue por necesidad lógica de la premisa, por ejemplo:

Todo el que es viudo fue casado

Todo el que es viudo no es mampostero.

El que todo viudo se casara una vez (lo cual es verdadero) no implica por necesidad (lógica o no) que ninguno de los viudos sea mampostero (lo cual es seguramente falso).

La noción fundamental en la formulación del segundo rasgo de la noción intuitiva de consecuencia lógica es la noción de expresión lógica. Pues el segundo rasgo es que todo argumento con la misma forma lógica que un argumento lógicamente correcto es también lógicamente correcto; y la forma lógica de un argumento es el resultado de sustituir de manera uniforme en él las expresiones no lógicas por letras esquemáticas. La noción de expresión lógica es nuevamente muy vaga, y la frontera entre expresiones lógicas y no lógicas, difusa. Pero a menudo las intuiciones que guían (o constituyen) el uso de esta noción son lo bastante claras como para determinar que ciertas expresiones particulares son lógicas y ciertas otras, no. Todas las expresiones que fueron sustituidas por letras esquemáticas en los argumentos de los ejemplos son intuitivamente no lógicas, y las que quedaron sin sustituir son intuitivamente lógicas.

¿Qué caracteriza a las expresiones no lógicas, o, lo que para el caso es lo mismo, a las lógicas? ¿Cómo se las puede arreglar uno para separarlas unas de otras de una manera fiable? Vamos a posponer el examen de estas cuestiones hasta la última sección de este artículo. Pero hasta llegar a esa sección daremos a menudo por supuesto que hay una división que parte el conjunto de las expresiones en dos conjuntos, el de las expresiones lógicas y el de las no lógicas.

Ahora que hemos explicado los dos rasgos intuitivos de la noción de consecuencia lógica podemos cerciorarnos de que argumentos particulares que son intuitivamente ejemplos de consecuencia lógica poseen efectivamente esos dos rasgos. Un argumento de una forma lógica familiar es éste:

Todo el que es sabio transige
Todo el que transige es pesimista
———————————————
Todo el que es sabio es pesimista.

Su conclusión se sigue por necesidad lógica de sus premisas. Y todo argumento de la misma forma lógica que éste es un argumento lógicamente correcto, de lo que uno puede convencerse si examina esa forma lógica, es decir,

Todo el que *F G*
Todo el que *G H*

Todo el que *F H*.

Otro argumento de la misma forma, que es por tanto lógicamente correcto, es

Todo el que es felino es mamífero
Todo el que es mamífero es vertebrado

Todo el que es felino es vertebrado.

Hemos dicho que la lógica se ocupa de caracterizar la noción de consecuencia lógica (y de buscar formas de separar de manera práctica los argumentos lógicamente correctos de los argumentos lógicamente incorrectos). Puede por tanto decirse que hemos hecho un poco de lógica, al menos porque hemos hecho algunos comentarios que probablemente han de ser útiles para caracterizar la noción de consecuencia lógica (y también porque hemos dado algunas sugerencias sobre cómo detectar la corrección e incorrección lógica de argumentos). Pero no hemos ofrecido una caracterización de la noción de consecuencia lógica en términos de condiciones necesarias y suficientes (y tampoco una manera muy buena de distinguir argumentos lógicamente correctos de argumentos lógicamente incorrectos).

Una caracterización que se le ocurre a uno inmediatamente a la vista de los comentarios precedentes es la siguiente: un argumento es un caso de consecuencia lógica si y sólo si su conclusión se sigue por necesidad lógica de sus premisas y todo argumento con la misma forma lógica es un caso de consecuencia lógica. Esta ocurrencia no sólo no es descabellada, sino que es trivialmente verdadera. La dirección «sólo si» es verdadera si aceptamos los dos rasgos intuitivos de la noción consecuencia lógica. La dirección «si» es verdadera, puesto que un argumento comparte forma lógica consigo mismo. Pero esta caracterización trivialmente verdadera deja mucho que desear. Nótese, sin ir más lejos, que es una caracterización circular, en la que el *definiendum* aparece en el *definiens*. Quizá no todos los círculos sean malos, pero éste lo es, y lo es por la misma razón que es mala otra caracterización, esta vez no circular, de la noción de consecuencia lógica.

Consideremos esta otra caracterización: un argumento es un caso de consecuencia lógica si y sólo si su conclusión se sigue por necesidad lógica de sus premisas y todo argumento con la misma forma ló-

gica es un argumento en el que la conclusión se sigue por necesidad lógica de las premisas. De nuevo la caracterización no es descabellada, y de hecho la dirección «sólo si» es verdadera. También es difícil imaginar que la dirección «si» sea falsa, es decir, que pueda haber un argumento lógicamente incorrecto donde la conclusión se sigue por necesidad lógica de las premisas y tal que todo argumento con la misma forma lógica es un argumento en el que la conclusión se sigue por necesidad lógica de las premisas. (Pero no afirmo categóricamente que esa dirección sea verdadera.)

Lo realmente malo de esta última caracterización es lo mismo que tiene de malo la caracterización circular anterior. Lo malo es que no explica o delimita la noción de consecuencia lógica en términos de nociones mejor comprendidas y más claras que ella. En el caso de la segunda caracterización, es muy cuestionable que nuestra comprensión del concepto de consecuencia por necesidad lógica sea superior a nuestra comprensión del concepto de consecuencia lógica. Naturalmente, ambos son conceptos altamente intuitivos que no se hallan respaldados por teorías o conjuntos de tesis bien establecidas más allá de las simples convicciones comunes sobre ellos. Ésta no sería una crítica pertinente a la segunda caracterización si el concepto de consecuencia por necesidad lógica fuese más claro o mejor comprendido que el de consecuencia lógica. El problema es que nuestra comprensión de ambos conceptos intuitivos parece *igualmente* pobre. Una consecuencia de esto es que, cuando consideramos la cuestión de si una cierta oración se sigue por necesidad lógica de un cierto conjunto de oraciones y todo argumento con la misma forma lógica es un argumento en el que la conclusión se sigue por necesidad lógica de las premisas, no se nos sugiere ninguna forma de decidirla que sea sustancialmente diferente de la que intuitivamente aplicaríamos si nos hiciéramos la pregunta de si el argumento correspondiente es un caso de consecuencia lógica.

En este artículo presentaremos, en primer lugar, una breve descripción de los dos tipos de aproximaciones históricamente más destacados al problema de caracterizar la noción de consecuencia lógica en términos de nociones mejor comprendidas que ella: el tipo de aproximaciones basadas en la noción de derivabilidad (sección 2) y el tipo de aproximaciones basadas en la noción de validez (sección 3) —con especial atención a una aproximación del segundo tipo, la tarskiana (sección 4)—. Luego ofreceremos un panorama de la discusión filosófica acerca del problema de si las caracterizaciones descritas son correctas (secciones 5 y 6). Por último, como ya adelantamos, expondremos y criticaremos algunos intentos de caracterizar la noción de forma lógica en términos de otras nociones (sección 7).

2. DERIVABILIDAD

Aristóteles, el fundador de la lógica como disciplina, ofreció una primera aproximación fructífera a la caracterización del concepto de consecuencia lógica. Aristóteles aisló un buen número de formas lógicas de argumentos y observó que los argumentos de unas cuantas formas lógicas básicas de las aisladas por él eran silogismos o argumentos lógicamente correctos. Además, observó también que la corrección lógica que se da en los argumentos de otras cuantas formas podía justificarse a partir de la de las formas básicas por aplicaciones sucesivas de reglas de transformación que intuitivamente producían argumentos lógicamente correctos cuando se aplicaban a argumentos lógicamente correctos. De este modo, Aristóteles caracterizó de una forma muy precisa un conjunto bastante amplio de argumentos lógicamente correctos: los argumentos de las formas identificadas por él como básicas más los obtenibles por sucesivas aplicaciones de sus reglas de transformación a partir de argumentos de las formas básicas. Los filósofos estoicos y megáricos aislaron muchas nuevas formas lógicas (que hoy se estudian en lógica proposicional o de enunciados), identificaron unas cuantas formas básicas lógicamente correctas y al parecer propusieron un cálculo de transformaciones inspirado en la misma idea que el de Aristóteles.

Frege, fundador de la lógica moderna a finales del siglo XIX, se aproximó a la relación de consecuencia lógica de un modo análogo, en esencia, al aristotélico, estoico y megárico: Frege identificó muy importantes nuevas formas lógicas y ofreció un cálculo de axiomas y reglas para delimitar (de la manera que se explica en el siguiente párrafo) un conjunto de formas lógicamente correctas. Pero muchas de sus contribuciones fueron sin embargo revolucionarias con respecto a la lógica anterior, y de tremenda profundidad. En primer lugar, Frege inventó un lenguaje formal (o una serie de lenguajes), diseñado especialmente para la formalización de argumentos matemáticos, dentro del cual siempre es enteramente claro cuál es la forma lógica de un argumento y si dos argumentos tienen la misma forma lógica o no. Frege conjeturó además que cuando menos todo argumento matemático que fuese lógicamente correcto podría formalizarse por medio de un argumento lógicamente correcto en su lenguaje formal[1]. El lengua-

1. No prestaremos atención en este artículo a la importante cuestión de si o en qué medida las formalizaciones de argumentos del lenguaje natural (como los matemáticos) en lenguajes formales como los de Frege reflejan plenamente todas las propiedades semánticas de esos argumentos. Es razonable suponer, sin embargo, que a menudo es posible construir un lenguaje formal con la propiedad de que los argumentos

je que inventó Frege era lo que hoy llamaríamos un lenguaje cuantificacional de orden superior. Este lenguaje contenía como fragmento lo que hoy llamaríamos un lenguaje cuantificacional de primer orden, un lenguaje como los que se presentan hoy en los cursos básicos de lógica. En segundo lugar, Frege incrementó de manera formidable el rigor de la presentación axiomática en un cálculo lógico, hasta el punto de que puede considerársele el creador de la idea de sistema formal: él fue el primero en ofrecer un sistema formal como los que se estudian hoy en los cursos de lógica.

Una vez construido un sistema así, es posible proponer una caracterización muy precisa del conjunto de argumentos lógicamente correctos del lenguaje formalizado del sistema: podemos proponer que la relación de consecuencia lógica se da entre un conjunto de premisas P_1, P_2, P_3, etc. y una conclusión C (del lenguaje del sistema) exactamente cuando existe una serie de aplicaciones de las reglas de inferencia que, partiendo de P_1, P_2, P_3, etc. y posiblemente también de oraciones de las formas axiomáticas básicas, acaba en C. Cuando una serie tal existe se dice que C es *derivable* en el sistema formal a partir de P_1, P_2, P_3, etc. Y ciertamente, como subrayaremos en la sección 5, si el sistema formal se construye con esmero, al concluir uno quedará convencido al menos de que todos los argumentos cuya conclusión es derivable de sus premisas (en el sistema) son argumentos lógicamente correctos, es decir, uno quedará convencido de que la derivabilidad de la conclusión a partir de las premisas (en el sistema) es una condición suficiente para que un argumento sea lógicamente correcto. La cuestión de si podemos convencernos de que es también una condición necesaria la trataremos en las secciones 5 y 6.

Nuestra comprensión de la relación de derivabilidad en un sistema como el de Frege es sin duda mejor y más clara que nuestra comprensión de los conceptos de consecuencia lógica y de consecuencia por necesidad lógica. La pregunta de si una conclusión es derivable en un sistema formal a partir de un conjunto de premisas es una pregunta esencialmente más precisa y mejor delimitada que las preguntas de si esa conclusión es implicada por necesidad lógica por las premisas, y de si es una consecuencia lógica de ellas. El acercamiento a la relación de consecuencia lógica en términos de la de derivabilidad en ciertos sistemas goza por tanto de un gran atractivo metodológico

lógicamente correctos de una cierta área del lenguaje natural son formalizables por medio de argumentos semánticamente muy similares y lógicamente correctos de ese lenguaje formal. Frecuentemente las afirmaciones de este artículo son acerca de argumentos de lenguajes formales, pero también frecuentemente hay versiones abstractas de esas afirmaciones que valen para argumentos del lenguaje natural.

y explicativo. En gran parte por estos motivos, este acercamiento proporcionaría la concepción dominante de la relación de consecuencia lógica entre los lógicos durante largo tiempo después de la obra de Frege. A lo largo del siglo XX se construyeron también sistemas formales para lenguajes formales más amplios que los de Frege (por ejemplo, para lenguajes con operadores modales, con cuantificadores ramificados, etc.), con el objetivo de aplicar las ventajas del método de Frege para identificar otras formas lógicas de argumentos lógicamente correctos.

3. VALIDEZ

En la lógica ha existido también desde el mismo Aristóteles un tipo de aproximación alternativo, y en cierto modo complementario, a la caracterización de la relación de consecuencia lógica. Este tipo de aproximación se basa plenamente en los dos rasgos intuitivos de la noción de consecuencia lógica. Recordemos que el segundo rasgo consiste en que todo argumento con la misma forma lógica que uno lógicamente correcto es también lógicamente correcto. Como señalamos, esto proporciona una condición necesaria de los argumentos lógicamente correctos, aunque en términos de la noción de corrección lógica (y también una condición necesaria igualmente poco iluminadora en términos de la noción de consecuencia por necesidad lógica). Pero también sugiere una condición necesaria en términos de la noción de verdad. Observemos que si un argumento es lógicamente correcto, entonces no tiene premisas verdaderas y conclusión falsa; pues si así fuera las premisas no implicarían por necesidad lógica la conclusión y entonces, por el primer rasgo de la noción de consecuencia lógica, el argumento no sería lógicamente correcto después de todo. Por tanto, por el segundo rasgo de la noción de consecuencia lógica, un argumento es lógicamente correcto sólo si ningún argumento con la misma forma lógica tiene premisas verdaderas y conclusión falsa. Ésta es la condición necesaria en términos de la noción de verdad a la que me refería antes.

El tipo de aproximación alternativo a la caracterización de la noción de consecuencia lógica usa siempre una condición que es una variante de esta condición necesaria, proponiéndola en cada caso como condición necesaria y suficiente. Las variantes de la condición ya mencionada surgen de la exigencia de que la condición propuesta sea efectivamente no sólo necesaria sino también suficiente, combinado con el requisito tácito de que la caracterización final se dé en térmi-

nos de nociones de una transparencia mayor que la de la noción misma de consecuencia lógica. La combinación de la exigencia y el requisito aparece de forma muy ilustrativa en la obra de Tarski, quien quizá sea el representante más paradigmático del tipo de aproximación alternativo. Examinemos brevemente la caracterización tarskiana de la noción de consecuencia lógica y la forma en que Tarski llega a ella. (En aras de la simplicidad, la exposición no será completamente rigurosa desde un punto de vista histórico; al lector que desee una explicación más exacta históricamente de las ideas de Tarski le recomiendo la lectura de mis trabajos [Gómez Torrente, 1996, 1998 y 2000a].)

En su conocido artículo «Sobre el concepto de consecuencia lógica» (Tarski, 1936), Tarski ofrece su caracterización con miras a su aplicación a los mismos lenguajes formales creados por Frege, y aceptando la noción de forma lógica para oraciones y argumentos de esos lenguajes implícita en Frege; estos lenguajes son, en esencia, los lenguajes que hoy llamamos cuantificacionales clásicos de órdenes primero y superiores. Sin embargo, es bueno subrayar que el método tarskiano se puede usar, y es el que de hecho se usa, para dar caracterizaciones similares de la noción de consecuencia lógica para lenguajes formales que amplían los lenguajes fregeanos.

Para propósitos ilustrativos será bueno tener en mente un lenguaje cuantificacional de primer orden. El lenguaje que escogemos será un lenguaje de primer orden para un fragmento muy sencillo de la aritmética elemental. Este lenguaje, *LAr*, tiene signos primitivos para el cuantificador universal de primer orden ('$\forall$'), el condicional ('$\rightarrow$'), la negación ('$\neg$'), paréntesis ('(', ')') (el cuantificador existencial de primer orden, '$\exists$', y las otras conectivas veritativo-funcionales, como '$\wedge$' y '$\vee$', se pueden definir de formas acostumbradas usando estos símbolos), igualdad ('='), la letra 'x', un acento subíndice ('$_{'}$') para generar una cantidad infinita numerable de variables por sufijación a 'x' (para abreviar, usaremos la notación 'x_n' para referirnos a la variable que consta de la letra 'x' seguida de *n* acentos subíndices), y el predicado diádico 'M', con el significado deseado «*a* es menor o igual que *b*». Suponiendo que las cuantificaciones hablan de números naturales, ejemplos de oraciones de este lenguaje y sus significados deseados son '$\exists x \forall x_{'} \neg M x_{'} x$' («existe un número natural x tal que para todo número natural $x_{'}$, $x_{'}$ no es menor o igual que x»), '$\forall x \exists x_{'} M x x_{'}$' («para todo número natural x existe un número natural $x_{'}$ tal que x es menor o igual que $x_{'}$»).

En las consideraciones preliminares a la propuesta de su caracterización, Tarski dice que cuando una oración *X* de un lenguaje formal como *LAr* es una consecuencia lógica de un conjunto *K* de oraciones

de ese lenguaje, el argumento con premisa *K* y conclusión *X* tiene la siguiente propiedad, que Tarski llama 'condición (F)':

> (F) Si, en las oraciones del conjunto *K* y en la oración *X*, las constantes no lógicas son sustituidas de manera uniforme por cualesquiera otras constantes no lógicas del mismo lenguaje, y si llamamos '*K*'' al conjunto así obtenido a partir de *K*, y '*X*'' a la oración obtenida a partir de *X*, entonces la oración *X*' es verdadera o alguna oración de *K*' es falsa (cf. Tarski, 1936, 415).

Aclaremos el sentido de la condición (F) por medio de un ejemplo. Consideremos un lenguaje LAr^{+} que es como *LAr* pero además tiene otro predicado diádico, 'P', cuyo significado deseado es «*a* es mayor que *b*». Sea *K* el siguiente conjunto unitario de oraciones de LAr^{+}: {'$\exists x \forall x_1 Mxx_1$'} (esta oración es verdadera); y sea *X* la oración '$\forall x \exists x_1 Mx_1x$'. El argumento con la premisa de *K* y conclusión *X* es intuitivamente correcto lógicamente, como muestra un razonamiento elemental (además, al ser su única premisa verdadera, su conclusión ha de ser verdadera, y en efecto podemos comprobar que lo es). Por tanto, según Tarski, ha de cumplir la condición (F). Esto quiere decir que *cualquier* argumento de LAr^{+} obtenido a partir de él por sustitución uniforme de constantes no lógicas por constantes no lógicas ha de ser un argumento donde no es el caso que las premisas sean verdaderas y la conclusión falsa. Supongamos que las constantes no lógicas de LAr^{+} son 'M' y 'P'. Sustituyamos 'M' por 'P' en el argumento con la premisa de *K* y conclusión *X* y llamemos al conjunto y la oración resultantes '*K*'' y '*X*''. Es decir, *K*' es {'$\exists x \forall x_1 Pxx_1$'} y *X*' es '$\forall x \exists x_1 Px_1x$'. En virtud de la condición (F), el argumento con premisas de *K*' y conclusión *X*' ha de ser un argumento donde no es el caso que las premisas sean verdaderas y la conclusión falsa; y en efecto es así: la conclusión es verdadera y la premisa falsa.

Tarski se pregunta si es posible ofrecer la condición (F) como caracterización de la relación de consecuencia lógica, es decir, si es aceptable proponer que la condición (F) es no sólo una condición necesaria sino también una condición suficiente para que un argumento sea un caso de consecuencia lógica. Su respuesta es que no es posible. La razón de esto es que la condición (F)

> puede satisfacerse de hecho meramente porque el lenguaje con el que nos las vemos no posee un acopio suficiente de constantes extra-lógicas. La condición (F) podría ser considerada como suficiente para que la oración *X* se siga del conjunto *K* sólo si los nombres de todos los objetos posibles aparecieran en el lenguaje en cuestión. Esta suposi-

> ción, sin embargo, es quimérica y nunca puede darse (Tarski, 1936, 415-416).

Es decir, Tarski observa que para que un argumento sea un caso de consecuencia lógica no tiene por qué ser suficiente que todos los argumentos de la misma forma lógica en el lenguaje en cuestión sean argumentos donde no es el caso que las premisas sean verdaderas y la conclusión falsa. Es concebible que sea posible interpretar las constantes no lógicas del argumento por medio de ciertos objetos (propiedades, relaciones, etc.) de manera que las premisas así reinterpretadas sean verdaderas y la conclusión así reinterpretada sea falsa, y que sin embargo (algunos de) esos objetos no estén denotados por constantes no lógicas del lenguaje que está siendo considerado; en tal caso no diríamos que el argumento es un caso de consecuencia lógica, a pesar de que satisfaría la condición (F).

Para poner un ejemplo, supongamos de nuevo que el lenguaje que estamos considerando es *LAr*$^{+}$. Puesto que tanto la relación «*a* es menor o igual que *b*» como la relación «*a* es mayor o igual que *b*» son transitivas en el dominio de los números naturales, la oración '$\forall x \forall x' \forall x''(Mxx' \land Mx'x'' \rightarrow Mxx'')$' sería declarada consecuencia lógica de todo conjunto de premisas por el criterio (F): ninguna sustitución de la constante no lógica 'M' por otras constantes no lógicas de *LAr*$^{+}$ convierte esa oración en otra falsa. Pero claramente '$\forall x \forall x' \forall x''(Mxx' \land Mx'x'' \rightarrow Mxx'')$' no es una consecuencia lógica de, digamos, '$\forall x x = x$'. Esto se puede justificar, por ejemplo, observando que 'M' puede interpretarse por medio de la relación intransitiva «*a* es el doble de *b*». Con esta interpretación, '$\forall x \forall x' \forall x''(Mxx' \land Mx'x'' \rightarrow Mxx'')$' es falsa, a pesar de que '$\forall x x = x$' es verdadera. (La observación de Tarski de que es quimérico suponer que todos los objetos tengan nombre en el lenguaje puede justificarse, por ejemplo, observando que hay una cantidad no numerable de relaciones de números naturales, pero en los lenguajes considerados por él sólo hay una cantidad numerable de expresiones.)

Debe notarse que la condición (F) es una variante de la condición que mencionamos antes, la condición de que *ningún* argumento con la misma forma lógica que uno dado tenga premisas verdaderas y conclusión falsa. En concreto, la condición (F) dice que ningún argumento con la misma forma lógica que uno dado *en el lenguaje del argumento dado* tiene premisas verdaderas y conclusión falsa. La diferencia esencial entre ambas condiciones es que en la condición (F) no se usa la noción de *argumento arbitrario*, sino la de argumento de un lenguaje previamente especificado. La razón de esta restricción tiene

que ver con el requisito tácito de que la caracterización final se dé en términos de nociones mejor comprendidas que la de consecuencia lógica. La condición no restringida sería objetable como caracterización por la simple razón de que la noción de argumento arbitrario es problemática para este propósito. ¿Qué es un argumento arbitrario? No es simplemente un argumento del español (por ejemplo). Tampoco está claro que sea un argumento de un lenguaje especificable aquí y ahora. (Por cierto, la objeción de Tarski a la condición (F) se aplicaría en estos casos también.) Al hablar de argumentos arbitrarios no estamos pensando simplemente en ristras de signos de un cierto tipo. Más bien estamos pensando en argumentos donde los términos tienen denotaciones posibles, o lógicamente posibles. Pero la noción de denotación lógicamente posible (como la de interpretación lógicamente posible, que veremos luego) no es más transparente que la de consecuencia lógica. Por tanto, la noción de argumento arbitrario no puede usarse en una caracterización de la noción de consecuencia lógica que sea más transparente que la noción común. Sin embargo, si usamos la condición transparente (F) (o una similar) nos encontramos con el problema señalado por Tarski.

La propuesta teórica de Tarski consistirá en ampliar el requisito expresado por la condición (F) de forma que se incorpore, aunque sea parcialmente, la idea de que un argumento lógicamente correcto no puede ser *reinterpretado* haciéndose verdaderas las premisas y falsa la conclusión (y no meramente la idea de que no puede ser *convertido* en un argumento con premisas verdaderas y conclusión falsa sustituyendo constantes por constantes en un lenguaje especificado). En otras palabras, la idea de que una oración *X* es consecuencia lógica de un conjunto de oraciones *K* cuando *toda interpretación en que todas las oraciones de K son verdaderas es una interpretación en que X es verdadera*. O, como a veces se dice, cuando toda interpretación *preserva* la verdad de las premisas en la conclusión. Cuando toda interpretación preserva la verdad de las premisas en la conclusión se dice también que el argumento es *válido*. Este calificativo se justifica por el hecho de que si un argumento es válido entonces claramente «vale», al menos en el sentido de que, incluso si no es lógicamente correcto, la verdad de su conclusión se puede concluir sin reparo de la de sus premisas.

La idea de Tarski es reflejar la condición de validez en su caracterización, y hacerlo de modo matemáticamente satisfactorio y manejable. Como dice Tarski, la idea de entender la noción de consecuencia lógica por medio de la noción de validez no es una idea original suya, sino implícita en la práctica lógica y matemática de su tiempo (espe-

cialmente la de los matemáticos y lógicos interesados en ofrecer pruebas de independencia). Pero lo novedoso es que Tarski propone hacer matemáticamente manejable esta idea empleando el aparato desarrollado por él para la definición matemática de conceptos semánticos, como los de satisfacción, definibilidad y verdad. Veamos cómo procede Tarski.

4. VALIDEZ TARSKIANA

Usemos el lenguaje *LAr* en nuestro ejemplo. Primero Tarski introduce la noción de *función oracional*. Una función oracional O' de una oración O es el resultado de sustituir las constantes no lógicas que aparecen en O de una manera uniforme por variables correspondientes de los tipos apropiados y diferentes de las variables ya existentes en el lenguaje. Por ejemplo, la función oracional determinada por la oración '∀xMxx' de *Lar* es la expresión '∀xΨxx' (en la cual 'Ψ' es una variable nueva asignada a 'M'). Igualmente es posible definir la noción más general de *función formular*, de forma análoga, salvo que en este caso la O de partida puede ser una fórmula abierta. Nótese que la función oracional a la que corresponde una oración puede verse como una representación de su forma lógica, o como su forma lógica a secas[2].

Las funciones oracionales de oraciones de *LAr* no serán en general oraciones, y por tanto no tendrán siempre la propiedad de ser verdaderas o falsas. Pero en cambio sí serán siempre verdaderas o falsas *con respecto a interpretaciones para LAr*. Estas interpretaciones son las cosas de las que se habla en la enunciación de la idea de validez. Pero ¿qué son? La idea es que las interpretaciones proporcionan denotaciones posibles (o lógicamente posibles) para las constantes no lógicas. Sin embargo, esta noción de interpretación lógicamente posible no es sustancialmente más transparente que la noción de implicación por necesidad lógica o la noción de necesidad lógica. Pues una interpretación lógicamente posible es una interpretación cuya inexistencia no es lógicamente necesaria.

¿Cómo hacer uso entonces, pero de una manera más transparente, de la noción de validez? La estrategia de Tarski consiste en seleccionar una clase amplia de interpretaciones lógicamente posibles que se pueda caracterizar de una manera aceptablemente transparente, y usarla de la manera obvia en la definición de una noción técnica de

2. El que sea una cosa u otra dependerá simplemente de nuestras convenciones acerca de qué letras se deben usar como letras esquemáticas.

validez para argumentos de lenguajes formales. La cuestión de si la clase de interpretaciones seleccionada por Tarski es suficientemente amplia, que se convertirá en la cuestión de si la condición de validez definida técnicamente por Tarski es suficiente para que un argumento sea lógicamente correcto, será analizada en las secciones 5 y 6.

Digamos que una *estructura para LAr* es una secuencia que tiene como primer miembro a un *conjunto*[3] no vacío *U* y que asigna un objeto apropiado «sacado» de *U* a la constante no lógica de *LAr* ('M'): una relación binaria *R* entre objetos de *U*. Definimos entonces el concepto de satisfacción de una función oracional por una estructura. Decimos que *la estructura* $<U, R>$ *satisface la función formular X con respecto a una secuencia f* (que asigne elementos de *U* a las variables de *LAr*) si y sólo si:

(α) (i) X es $\Psi x_n x_m$ (para algunos m y n) y $<f(x_n),f(x_m)>\in R$; o
(ii) X es $x_n = x_m$ (para algunos m y n) y $f(x_n) = f(x_m)$; o

(β) hay una función formular Y tal que X es $\neg Y$ y $<U, R>$ no satisface Y con respecto a la secuencia f; o

(γ) hay funciones formulares Y y Z tales que X es $(Y\rightarrow Z)$ y o bien $<U, R>$ no satisface Y con respecto a la secuencia f o $<U, R>$ satisface Z con respecto a la secuencia f; o, por último,

(δ) hay una función formular Y y un número n tales que X es $\forall x_n Y$ y toda secuencia g que asigna valores en U a las variables de *LAr* y que difiere de f a lo sumo en lo que asigna a x_n es tal que $<U, R>$ satisface Y con respecto a g.

3. Nótese que las estructuras para *LAr* proporcionan no sólo una denotación para la constante no lógica de *LAr* sino también un universo *U* de objetos del cual se saca esa denotación y que proporciona el recorrido de la cuantificación. La razón es que es muy natural suponer que los cuantificadores de un lenguaje formal, así como los cuantificadores empleados en un contexto de uso de un fragmento del lenguaje natural, hablan a menudo de los objetos de un dominio específico (las personas, los animales, los números naturales —para mencionar los dominios de los que se habla en varios de los ejemplos de este artículo—) y no de objetos arbitrarios (actuales, posibles, lógicamente posibles...), aunque también puedan usarse para hablar de objetos arbitrarios. Si se adopta esa suposición es claro que una interpretación que asigne un valor de verdad a una oración con cuantificadores no puede consistir simplemente en una lista de denotaciones para las expresiones no lógicas. Además, es claro que la corrección lógica de un argumento no debe depender de las peculiaridades del dominio de cuantificación específico de su lenguaje, y que si un argumento es lógicamente correcto, entonces el hecho de que sea correcto debe tener aplicación en el razonamiento sobre cualquier otro dominio y no sólo el específico de su lenguaje. (Sobre el requisito de que el universo de una estructura no sea vacío volveremos en una nota posterior.)

Es ésta una definición recursiva, enteramente paralela a la definición de satisfacción de fórmulas por secuencias en el artículo de Tarski «El concepto de verdad en los lenguajes formalizados» (Tarski, 1935, 193; he procurado usar la misma terminología que en este artículo clásico de Tarski). De la misma manera que esa definición, la que acabamos de dar puede hacerse también explícita de forma enteramente análoga.

Podemos decir entonces que *la estructura* <*U*, *R*> *satisface la función oracional X* si y sólo si <*U*, *R*> satisface la función formular *X* con respecto a toda secuencia *f* que asigna valores en *U* a las variables de *LAr* (una definición análoga a la definición de verdad de Tarski). Decimos también que O *es verdadera en* una estructura <*U*, *R*> si <*U*, *R*> satisface la función oracional O' determinada por O. De forma más general, un conjunto de oraciones *K* es verdadero en una estructura si es una estructura que satisface todas las funciones oracionales determinadas por oraciones de *K*. Finalmente, en términos de este aparato de nociones podemos ofrecer la condición por medio de la que Tarski caracteriza la noción de consecuencia lógica para argumentos con premisas de *K* y conclusión O:

(VT) Toda estructura en la que el conjunto *K* es verdadero es también una estructura en la que la oración O es verdadera. (Abreviaremos 'Toda estructura en la que el conjunto *K* es verdadero es también una estructura en la que la oración O es verdadera' por medio de la notación 'Val_T(*K*,O)'.)

'VT' está por «validez tarskiana». Nociones análogas de validez tarskiana se pueden definir para otros lenguajes formales usando el mismo método que hemos seguido para *LAr* (incluidos lenguajes que amplían a los fregeanos y para los cuales se haya definido previamente una noción razonable de estructura).

El subíndice 'T' lo uso para poner énfasis en el hecho de que 'Val_T(*K*,O)' denota una relación distinta de la de validez a secas. Recuérdese que los dominios de las estructuras para *LAr* son *conjuntos*. Aquí la palabra 'conjunto' se está usando en el sentido que tiene en matemáticas. Esto convierte a la propiedad tarskiana Val_T(*K*,O) en una propiedad matemáticamente manejable y aceptablemente transparente para el propósito de caracterizar la noción de consecuencia lógica. En esto se diferencia de la relación de validez a secas. La relación de validez a secas, recuérdese, es la relación que se da entre un conjunto de premisas *K* y una oración O cuando toda *interpretación* en que todas las oraciones de *K* son verdaderas es una interpretación en que O es verdadera.

5. CORRECCIÓN Y COMPLECIÓN

Como señalamos en la sección 2, si uno construye un sistema formal con cuidado, podrá convencerse de que todos los argumentos cuya conclusión es derivable de sus premisas son argumentos lógicamente correctos. La razón de esto es que uno puede usar su intuición de un modo muy sistemático para obtener ese convencimiento: uno puede incluir en su sistema axiomas que le parezcan a uno consecuencias lógicas de cualquier conjunto de premisas; y uno puede incluir como reglas de derivación de su sistema reglas que le parezca a uno que producen oraciones que se siguen lógicamente de las oraciones a las que se aplican. Entonces, dada la definición de derivabilidad para el sistema formal que vimos en la sección 2, es inmediatamente claro para uno que no se podrá derivar de un conjunto de oraciones ninguna oración que no se siga lógicamente de ese conjunto de oraciones. Empleando otra terminología, que se explica a sí misma, podemos decir que, si uno construye su sistema formal con cuidado, la caracterización correspondiente en términos de derivabilidad (en ese sistema) es *correcta* con respecto a la noción de consecuencia lógica —o, simplemente, que la derivabilidad es correcta con respecto a la consecuencia lógica.

De igual modo, es intuitivamente obvio que si uno tiene a mano una noción tarskiana de validez para un lenguaje dado, entonces todos los argumentos lógicamente correctos (del lenguaje) serán argumentos válidos en el sentido tarskiano. La razón es simple: si un argumento no es válido en el sentido tarskiano, entonces hay una estructura, y por tanto una interpretación, que hace verdaderas sus premisas y falsa su conclusión. Por tanto sería en principio posible construir un argumento de la misma forma lógica, cuyos términos tendrían denotaciones lógicamente posibles, y que tendría premisas verdaderas y conclusión falsa. Pero el segundo rasgo intuitivo de la noción de consecuencia lógica implica que si el argumento original era lógicamente correcto entonces no hay ningún argumento de la misma forma lógica con premisas verdaderas y conclusión falsa. Habiendo concluido que todos los argumentos lógicamente correctos son válidos en el sentido tarskiano, podemos decir, empleando otra terminología, que la caracterización en términos de validez tarskiana es *completa* con respecto a la noción de consecuencia lógica —o, simplemente, que la validez tarskiana es completa con respecto a la consecuencia lógica.

Usemos las siguientes dos abreviaturas: '$\mathrm{Der}_S(K,O)$' para '*O* es derivable de *K* en el sistema formal *S*', y 'CL(*K*,*O*)' para '*O* es consecuencia lógica (en el sentido intuitivo) de *K*'. Entonces, si *S* es un sis-

tema formal construido con cuidado, la situación a la que hemos llegado se resume en el siguiente diagrama:

$$(1)\ \mathrm{Der}_S(K,O) \Rightarrow \mathrm{CL}(K,O) \Rightarrow \mathrm{Val}_T(K,O).$$

La primera implicación es la corrección de la derivabilidad con respecto a la consecuencia lógica; la segunda implicación es la compleción de la validez tarskiana con respecto a la consecuencia lógica. Ahora bien, para convencernos de que las caracterizaciones de la noción de consecuencia lógica en términos de $\mathrm{Der}_S(K,O)$ y $\mathrm{Val}_T(K,O)$ son apropiadas tendríamos que convencernos también de las implicaciones conversas:

$$(2)\ \mathrm{Val}_T(K,O) \overset{?}{\Rightarrow} \mathrm{CL}(K,O) \overset{?}{\Rightarrow} \mathrm{Der}_S(K,O),$$

o sea, de que la validez tarskiana es *correcta* con respecto a la consecuencia lógica, y de que la derivabilidad es *completa* con respecto a la consecuencia lógica. Convencerse de que esto es el caso, o de que no es el caso, resulta ser una tarea difícil (quizá sorprendentemente difícil) para un buen número de lenguajes, y examinaremos las dificultades de un tipo particular de lenguajes en la sección 6. Pero antes veremos una manera de convencerse, en algunos casos, de que las implicaciones con interrogante se dan de hecho. Esa manera de convencerse se basa en una observación sencilla pero profunda de Kreisel (1967).

Hay una cantidad considerable de lenguajes formales para los que existen nociones de validez tarskiana y de derivabilidad en un sistema *S*. Entre éstos, hay un buen número para los que la validez tarskiana es intuitivamente completa y la derivabilidad correcta con respecto a la consecuencia lógica. (No es así en todos los casos, pues a menudo se definen nociones de validez tarskiana o de derivabilidad con las que no se pretende caracterizar la noción de consecuencia lógica del lenguaje de turno, sino alguna otra noción, por ejemplo la noción de implicación analítica.) En estos últimos casos se dan las implicaciones de (1). Y a su vez, entre estos últimos lenguajes, hay muchos para los que es posible dar una demostración matemática de que la derivabilidad es completa con respecto a la validez tarskiana, o sea, una demostración de esta otra implicación:

$$(3)\ \mathrm{Val}_T(K,O) \Rightarrow \mathrm{Der}_S(K,O).$$

Kreisel llamó la atención sobre el hecho de que (3) (junto con (1)) implica que la validez tarskiana es correcta con respecto a la consecuencia lógica, o sea, que se da la primera implicación de (2)[4]. Esto quiere decir que cuando se da (3) la noción de validez tarskiana ofrece una caracterización apropiada de la de consecuencia lógica. Puede añadirse a lo subrayado por Kreisel que (3) (junto con (1)) implica que la derivabilidad en *S* es completa con respecto a la consecuencia lógica, o sea, que se da la segunda implicación de (2). Esto quiere decir que cuando se da (3) la noción de derivabilidad en un cierto sistema ofrece también una caracterización apropiada de la noción de consecuencia lógica.

Un caso especialmente significativo en que se da la implicación (3) (para ciertos sistemas formales *S*) y las implicaciones (1), y por tanto se dan las implicaciones (2), es el de los lenguajes cuantificacionales de primer orden[5]. Eso quiere decir que uno puede convencerse de que tanto la noción de derivabilidad como la de validez tarskiana (definidas de un modo apropiado para esos lenguajes) son caracterizaciones apropiadas —que no se dejan nada bueno fuera ni meten nada malo dentro— de la noción intuitiva de consecuencia lógica

4. Estrictamente hablando, algo más cercano a lo que señaló Kreisel es que la compleción de la derivabilidad en un sistema formal con respecto a una noción de validez tarskiana implica que esta última noción es correcta con respecto a la de validez a secas (si uno se ha convencido previamente de que la derivabilidad es correcta con respecto a la validez a secas, lo cual es factible por el mismo método por el que uno se puede convencer de que la derivabilidad es correcta con respecto a la consecuencia lógica).

5. Será bueno hacer un comentario a esta afirmación. Recuérdese que se pidió que el universo de una estructura para *LAr* fuese *no vacío*. Ésta es la práctica común cuando se define una noción técnica de estructura para lenguajes cuantificacionales. Sin embargo, esta práctica tiene como consecuencia que fórmulas como '$\exists x \exists x_1 (Mxx_1 \rightarrow Mxx_1)$' resultan conclusiones tarskianamente válidas de todo conjunto de premisas, lo cual choca con la intuición de muchos —«una afirmación existencial no puede ser lógicamente necesaria»—. El motivo principal de esa práctica de los lógicos es de orden pragmático. La colección de las interpretaciones conjuntistas con universo no vacío contiene todas las interpretaciones interesantes consideradas por los lógicos y matemáticos, y las oraciones verdaderas en todas las interpretaciones de esa colección forman un conjunto interesante, algo más amplio que el de las oraciones verdaderas en todas las interpretaciones conjuntistas. Además, la construcción de sistemas formales se simplifica notablemente con la suposición de que es permisible introducir en las derivaciones fórmulas verdaderas en toda estructura con universo no vacío pero no en toda interpretación conjuntista. Pero no entraña dificultad definir una noción de validez tarskiana para interpretaciones conjuntistas en general (con universo vacío y no vacío), y tampoco es difícil construir sistemas formales que cumplen (3) (y (1)) para esa noción de validez tarskiana. (Cuando uno tiene en mente una versión de los lenguajes de primer orden con constantes individuales y signos de función, un paso previo es adoptar una convención habitual para eliminarlos por medio de constantes predicativas y expresiones lógicas.)

para los lenguajes de primer orden. La situación no es tan clara en otros lenguajes especialmente significativos para la tradición lógica, los cuantificacionales de órdenes superiores a 1. A un comentario de esta situación está dedicada la sección siguiente.

6. CONSECUENCIA LÓGICA Y LENGUAJES DE ORDEN SUPERIOR

Los lenguajes cuantificacionales de orden superior fueron creados por Frege para formalizar un cierto tipo de afirmaciones matemáticas en las que se habla de «todas las propiedades», «todas las clases», «todas las relaciones diádicas», etc., de objetos de un conjunto dado (objetos que pueden ser propiedades o clases ellos mismos), por ejemplo todas las propiedades de números naturales, todas las clases de números reales, etc.[6]. Esas cuantificaciones se formalizan añadiendo variables nuevas a un lenguaje de primer orden donde las cuantificaciones son sobre el conjunto que interesa. En el caso de los lenguajes de segundo orden, uno añade variables de segundo orden a un lenguaje como *LAr*: las variables monádicas X^{1}_{I}, X^{1}_{II}, X^{1}_{III}, etc. (que intuitivamente representan propiedades), las variables diádicas X^{1}_{I}, X^{1}_{II}, X^{1}_{III}, etc. (que intuitivamente representan relaciones diádicas), las variables triádicas, etc. Una cuantificación universal de segundo orden es entonces una expresión del tipo $\forall X^{n}_{m}$. No es infrecuente entre los lógicos considerar que las cuantificaciones de orden superior son expresiones lógicas, y por tanto que forman parte de la forma lógica de los argumentos que las contienen (aunque no es una idea universalmente compartida, en parte por problemas como los que vamos a ver).

Hay también una manera muy natural de definir una noción de validez tarskiana para los lenguajes de orden superior, que se basa fundamentalmente en añadir una nueva condición a la definición de satisfacción de una función formular (ahora de orden superior) por una estructura, ahora con respecto a una secuencia *f* que asigne elementos de *U* a las variables de primer orden y a una secuencia *g* que asigne valores apropiados sacados de *U* a las variables de orden superior. He aquí el resultado para la ampliación de segundo orden de *LAr* —a la que podemos llamar *LAr*^{2}—. Decimos que *la estructura <U, R> satisface la función formular X con respecto a una secuencia f* (que asigne elementos de *U* a las variables de primer orden) *y a una secuencia g* (que asigne un subconjunto de U^k a cada variable *k*-ádica de segundo orden) si y sólo si:

6. Recuérdense el axioma de inducción de la aritmética de los números naturales y el axioma de compleción de la aritmética de los números reales.

(α) (i) X es $\Psi x_n x_m$ (para algunos m y n) y $<f(x_n),f(x_m)> \in R$; o
(ii) X es $x_n = x_m$ (para algunos m y n) y $f(x_n) = f(x_m)$; o
(iii) X es $X^n_m x_{i_1},...,x_{i_n}$ (para algunos m, n, $i_1,...,i_n$) y $<f(x_{i_1}),...,f(x_{i_n})> \in g(X^n_m)$; o

(β) hay una función formular Y tal que X es $\neg Y$ y $<U, R>$ no satisface Y con respecto a f y g; o

(γ) hay funciones formulares Y y Z tales que X es $(Y \rightarrow Z)$ y o bien $<U, R>$ no satisface Y con respecto a f y g o $<U, R>$ satisface Z con respecto a f y g; o

(δ) hay una función formular Y y un número n tales que X es $\forall x_n Y$ y toda secuencia h que asigna valores en U a las variables de primer orden y que difiere de f a lo sumo en lo que asigna a x_n es tal que $<U, R>$ satisface Y con respecto a h y g; o, por último,

(ε) hay una función formular Y y números n y m tales que X es $\forall X^n_m Y$ y toda secuencia h que asigna un subconjunto de U^k a cada variable k-ádica de segundo orden y que difiere de g a lo sumo en lo que asigna a X^n_m es tal que $<U, R>$ satisface Y con respecto a f y h.

Partiendo de esta definición, y siguiendo una cadena de definiciones enteramente análogas a las del caso de primer orden, se define entonces la noción de validez tarskiana para LAr^2. También el caso de órdenes superiores a 2 es análogo.

Sin embargo, es posible demostrar que para un lenguaje de segundo orden como LAr^2 *no existe ningún sistema formal*[7] *S* que cumpla las implicaciones (1) y que haga verdadera la implicación (3) para la noción de validez tarskiana correspondiente a ese lenguaje. Podemos llamar a este resultado *la incompleción de los sistemas formales de segundo orden con respecto a la validez tarskiana*. De hecho se da un resultado más fuerte, según el cual no hay ningún conjunto de oraciones para el que exista un sistema formal que cumpla (1) y que permita derivar del conjunto todas sus consecuencias válidas en el sentido tarskiano; dicho de otra manera, para todo conjunto de oraciones J y para todo sistema S que cumpla (1) existirá una oración O tal que $Val_T(J,O)$ pero no es el caso que $Der_S(J,O)$. Podemos llamar a este resultado *la incompleción fuerte de los sistemas formales de segundo orden con respecto a la validez tarskiana*.

7. En el sentido técnico de 'sistema formal' que se desprende de alguna definición intuitivamente adecuada de la noción de enumerabilidad efectiva, por ejemplo en términos de funciones recursivas.

En esta situación no es posible aplicar el argumento de Kreisel para convencerse de que sea apropiada la caracterización de la noción de consecuencia lógica para lenguajes de orden superior por medio de la noción de validez tarskiana: aunque se cumplen las implicaciones (1), no podemos aplicar el argumento de Kreisel para convencernos de que se cumplen las implicaciones (2). De hecho, la incompleción de los sistemas de segundo orden con respecto a la validez tarskiana, junto con (1), muestra que, dado cualquier sistema formal *S*, una de las implicaciones de (2) es falsa (o *ambas* lo son): *o* la derivabilidad en *S* es incompleta con respecto a la consecuencia lógica *o* la validez tarskiana es incorrecta con respecto a la consecuencia lógica. Distintos autores han extraído lecciones opuestas de esta situación. Explicaremos brevemente unas cuantas reacciones en ambos sentidos.

Una reacción muy común (aunque quizá igual de común que la reacción opuesta) es pensar que ha de ser la noción de validez tarskiana la que es incorrecta con respecto a la de consecuencia lógica. Una idea intuitiva en la que se basan a veces los que reaccionan así es la idea de que los argumentos lógicamente correctos, o, como mínimo, los argumentos lógicamente correctos con un número finito de premisas, deben ser argumentos en los que la conclusión sea extraíble de las premisas por medio de un *razonamiento*, probablemente incluso un razonamiento *a priori*. A esta idea se añade a veces la premisa de que el razonamiento (*a priori*) ha de ser reproducible de forma efectiva, en un sistema formal, si es que ha de ser posible que lo efectúen seres que razonen (*a priori*). De la idea y la premisa, junto a la incompleción fuerte de los sistemas de segundo orden con respecto a la validez tarskiana, se saca la conclusión de que ha de haber argumentos con un número finito de premisas que son válidos en el sentido tarskiano pero no son argumentos justificables por razonamiento (*a priori*).

Esta manera de argumentar, a pesar de que es relativamente frecuente, es claramente incorrecta. Supongamos que aceptamos que (i) el razonamiento (*a priori*) ha de ser reproducible por medio de un sistema formal. Aceptamos también, naturalmente, que (ii) para todo conjunto finito de premisas y *para todo* sistema formal que cumpla (1) *existen* argumentos válidos en el sentido tarskiano con esas premisas y cuya conclusión no es derivable de las premisas en el sistema formal. Pero de esto no se sigue que (iii) *existen* argumentos válidos en el sentido tarskiano con esas premisas y tales que *para todo* sistema formal que cumpla (1), su conclusión no es derivable de sus premisas en ese sistema. De (iii) y (i) se sigue desde luego que hay argumentos válidos en el sentido tarskiano que no son justificables por razonamiento (*a priori*). Pero el paso de (ii) a (iii) es incorrecto (se basa en un

error típico producido por la ambigüedad de ciertos enunciados con varios cuantificadores en el lenguaje natural, que se aclara en la formalización distinguiendo entre maneras de ordenar los cuantificadores). De (i) y (ii) no se sigue que exista ningún argumento válido en el sentido tarskiano que no sea justificable por razonamiento (*a priori*).

Otra idea intuitiva en la que se basan a veces los que reaccionan en contra de la corrección de la validez tarskiana es la idea de que los argumentos lógicamente correctos deben ser argumentos en los que la conclusión se sigue de las premisas *por implicación analítica*, es decir, en virtud del significado de las expresiones del argumento. Basándose en esta idea, John Etchemendy (1990) propuso el siguiente argumento en contra de la corrección de la validez tarskiana con respecto a la consecuencia lógica. La primera premisa del argumento es que hay dos oraciones de segundo orden *A* y *B* tales que

(a) $\mathrm{Val}_T(\varnothing, A)$ si y sólo si HC;
(b) $\mathrm{Val}_T(\varnothing, B)$ si y sólo si ¬HC.

Aquí 'HC' está en lugar de una formulación natural de una proposición algo complicada enunciable en el lenguaje de la teoría de conjuntos, la «hipótesis del continuo», que dice que no hay conjuntos de cardinalidad intermedia entre la del conjunto de los números naturales y la del conjunto de los conjuntos de números naturales. Las equivalencias (a) y (b) son en efecto verdaderas, y demostrables en la teoría de conjuntos. El hecho de que HC y su negación son complicadas lleva a nuevas premisas del argumento de Etchemendy:

(c) la hipótesis del continuo no es analítica;
(d) la negación de la hipótesis del continuo no es analítica

(donde 'analítica' significa «verdadera en virtud del significado de sus expresiones»). Por último, la premisa final es

(e) para toda oración O, si no es analíticamente verdadero que $\mathrm{Val}_T(\varnothing, O)$, entonces O no es analítica.

De (a) y (c) Etchemendy concluye

(f) no es analíticamente verdadero que $\mathrm{Val}_T(\varnothing, A)$;

y de (b) y (d) concluye

(g) no es analíticamente verdadero que $\mathrm{Val}_T(\varnothing, B)$.

De (e), (f) y (g) se sigue que ni *A* ni *B* son analíticas. Sin embargo, o $\mathrm{Val}_T(\varnothing,A)$ o $\mathrm{Val}_T(\varnothing,B)$, ya que o HC es verdadera o ¬HC lo es. Así pues, concluye el argumento, hay oraciones que son consecuencias válidas en el sentido tarskiano del conjunto vacío de premisas[8] y que no son analíticas. Dado el supuesto razonable de que una oración es analítica si y sólo si es una consecuencia por implicación analítica del conjunto vacío de premisas, de esa conclusión se sigue que hay argumentos que son casos de validez tarskiana pero no de consecuencia por implicación analítica.

Un primer elemento sospechoso en el argumento de Etchemendy es que (f) y (g) *no se siguen* de (a), (b), (c) y (d). Para concluir de «*p* si y sólo si *q*» y «*q* no es analíticamente verdadero» que «*p* no es analíticamente verdadero», la equivalencia entre *p* y *q* afirmada por «*p* si y sólo si *q*» ha de ser una equivalencia analítica, no una equivalencia puramente material o a lo sumo matemáticamente necesaria, como las que se afirman en (a) y (b). Pero aun suponiendo que concedamos (f) y (g) sin exigir un argumento, un elemento aún más sospechoso es la aceptación de la tesis (e). Etchemendy no ofrece *ninguna* justificación para (e), y no veo razón para pensar que sea verdadera. Etchemendy acepta (e), en mi opinión, porque en el texto en que la introduce la deriva falazmente de una tesis más plausible (y en la que Etchemendy ha insistido mucho en su obra, aunque innecesariamente):

(h) para toda oración O, no es analíticamente verdadero que [si $\mathrm{Val}_T(\varnothing,\mathrm{O})$ entonces O es analítica].

(h) es ciertamente muy plausible, y seguramente nadie ha pretendido nunca que sea falsa. Pero pasar de (h) a (e) es cometer una falacia modal muy típica y muy dañina. (El lector encontrará un examen más detenido de este asunto en Gómez Torrente [1998-1999].)

En general, no hay argumentos filosóficamente satisfactorios para pensar que la validez tarskiana sea incorrecta con respecto a la consecuencia lógica en lenguajes de orden superior[9]. ¿Hay buenas razo-

8. Nótese, por cierto, que las oraciones *O* tales que $\mathrm{Val}_T(\varnothing,O)$ son las mismas que las oraciones O tales que, para todo conjunto de oraciones *K*, $\mathrm{Val}_T(K,O)$.

9. Hay en la bibliografía argumentos de otro tipo que los que hemos visto hasta ahora para defender la incorrección de la validez tarskiana. En McGee (1992) hay un buen ejemplo. La idea de los que proponen estos argumentos es encontrar una oración O tal que $\mathrm{Val}_T(\varnothing,\mathrm{O})$ pero que intuitivamente sea falsa en una interpretación no conjuntista (por ejemplo, en una clase propia). Un análisis crítico y algo técnico de algunos argumentos de este tipo, que concluye que son insatisfactorios, puede verse en Gómez To-

nes entonces para pensar que la derivabilidad (en cualquier sistema que cumpla (1)) ha de ser incompleta con respecto a la consecuencia lógica? Tampoco creo que haya razones absolutamente convincentes. Examinemos un argumento en defensa de esa incompleción, con el cual ilustraremos una dificultad fundamental.

El argumento aparece en Tarski (1936). Tarski observa que en ciertos lenguajes de orden superior, por ejemplo en el lenguaje de una teoría simple de los tipos finitos, es posible formalizar argumentos de la forma siguiente, sin usar numerales como primitivos ni tomar el predicado 'número natural' como primitivo, sino usando exclusivamente expresiones lógicas de orden superior y un predicado P cualquiera:

$$\frac{\begin{array}{l} P(0), \\ P(1), \\ P(2), \\ \ldots \end{array}}{\text{Para todo número natural } n, P(n)}$$

(en lugar de los puntos suspensivos deberían ir el resto de las oraciones de la forma $P(n)$, para todo número natural $n>2$). Podemos llamar a estos argumentos *argumentos-ω*. Ahora bien, lo siguiente es una consecuencia relativamente fácil del primer teorema de incompleción de Gödel: para todo sistema formal del lenguaje de la teoría simple de los tipos finitos que cumpla (1) existirán argumentos-ω tales que la conclusión no es derivable de las premisas en el sistema formal. Podemos llamar a este resultado *la incompleción-ω de los sistemas formales del lenguaje de tipos finitos.*

Tarski subraya el hecho de que los argumentos-ω son claramente válidos, en el sentido de que no hay ni puede haber una interpretación de P que haga verdaderas las premisas y falsa la conclusión. Y ciertamente la verdad de la conclusión se puede concluir sin reparo de la de las premisas; la validez de estos argumentos es muy transparente, en un sentido en que no es transparente, por ejemplo, cuál de las oraciones A y B mencionadas por Etchemendy es una consecuencia válida en el sentido tarskiano del conjunto vacío de premisas. Quizá pueda decirse incluso que los argumentos-ω son casos de implicación por necesidad lógica, y aun casos de implicación analítica (aunque se-

rrente (1998-1999). Pero notemos que este tipo de argumentos ofrecen sólo un reto a la corrección de la caracterización de la consecuencia lógica en términos de estructuras *conjuntistas*, pero no un reto general a las caracterizaciones basadas en la idea de *validez*.

guramente no de implicación *a priori*, pues tienen un número infinito de premisas).

De la incompleción-ω Tarski concluye que la derivabilidad (en cualquier sistema formal) ha de ser incompleta con respecto a la consecuencia lógica —y eso le lleva a proponer su propia caracterización en términos de la noción de validez tarskiana—. Pero no es la de Tarski una conclusión absolutamente indiscutible. El problema fundamental es que esa conclusión se basa en el supuesto, que no hemos examinado críticamente, de que las expresiones típicamente catalogadas como lógicas en los lenguajes de orden superior, y en particular las cuantificaciones del tipo ∀X (donde X es una variable de orden superior), son efectivamente expresiones lógicas. Si lo son, entonces la observación de Tarski puede convencernos de que a todo sistema formal que cumpla (1) se le «escaparán» argumentos que son casos de implicación por necesidad lógica, y tales que todos los argumentos de la misma forma lógica son casos de implicación por necesidad lógica. Y esto ciertamente hace muy razonable pensar que si las cuantificaciones del tipo ∀X son expresiones lógicas, entonces la derivabilidad (en cualquier sistema formal que cumpla (1)) ha de ser incompleta con respecto a la consecuencia lógica[10]. Pero a falta de consideraciones adicionales, un crítico puede negar el supuesto en que se basa este argumento, y esto lo hace inconcluyente.

Esta última cuestión nos lleva directamente al último tema que trataremos en este artículo. ¿Es posible dar una caracterización intuitivamente apropiada, transparente e iluminadora de la noción intuitiva de expresión lógica? Y, más específicamente, ¿es posible dar una caracterización intuitivamente apropiada que dé una respuesta más concluyente o satisfactoria a disputas como la que acabamos de comentar?

7. EXPRESIONES LÓGICAS

No son preguntas fáciles de contestar, aunque hay un buen número de indicios que favorecen el pesimismo, o sea, que favorecen una respuesta negativa. En mi opinión, la noción de expresión lógica es considerablemente más vaga e intransparente que la de consecuencia por

10. Debe señalarse, sin embargo, que Tarski no hace uso en sus consideraciones de la noción de «implicación por necesidad lógica». Algo más próximo a lo que él tiene en mente es un razonamiento en el que la corrección lógica de un argumento se infiere de su validez a secas. (Como hemos señalado, no hay duda de que los argumentos-ω son válidos a secas.)

necesidad lógica —que ya es decir—. Parte de lo que la hace vaga e intransparente es el hecho de que los principios que han guiado (tanto tácita como explícitamente) a los lógicos al catalogar ciertas expresiones como lógicas han sido principios esencialmente pragmáticos con un contenido altamente vago en sí mismos. Típicamente, se ha visto a la lógica como una disciplina que trata el razonamiento general, el razonamiento aplicable en todos los campos argumentativos[11]. Por tanto es razonable pensar que un principio que subyace en la catalogación de expresiones como lógicas es el principio de que la lógica debe ocuparse de los argumentos correctos en virtud de las propiedades de las expresiones que se emplean de manera general en el razonamiento, expresiones no específicas de ninguno de los campos argumentativos sino comunes a todos ellos.

Este principio sugiere sólo una condición necesaria de las expresiones lógicas. Sin duda hay otros principios implícitos en juego. Preposiciones como 'para', 'en' y 'con' y verbos como 'añadir', 'incluir' y 'excluir' (en algunas acepciones) son expresiones que se emplean generalmente en el razonamiento y que no son específicas de ningún campo argumentativo; pero seguramente la intuición generalizada es que no cuentan ni deberían contar como expresiones lógicas. Hay cuestiones de grado y utilidad específica en juego. Los lógicos seguramente requieren implícitamente que las expresiones lógicas sean *muy relevantes* en el razonamiento en general, o que su estudio sea *útil* para la solución de problemas particularmente significativos en los razonamientos, para mencionar sólo dos posibilidades. Si ésta es la concepción correcta, entonces hay una gran vaguedad y al mismo tiempo complejidad en el concepto intuitivo de expresión lógica. Los principios pragmáticos que subyacen en este concepto dejan mucho (quizá demasiado) espacio a las interpretaciones divergentes de esos principios y, consiguientemente, a ideas incompatibles sobre *qué expresiones* son lógicas. De todos modos, esto no quiere decir que sea arbitrario qué expresiones son lógicas, pues los principios mencionados no son compatibles con cualquier idea acerca de qué expresiones son lógicas.

11. Así, Aristóteles dice: «Todas las ciencias se asocian unas con otras en las cosas comunes (llamo comunes a las que usan como puntos de partida en sus demostraciones —no a aquellas acerca de las que demuestran algo, ni a aquellas que son demostradas—); y la lógica se asocia con todas ellas, como ciencia que intenta probar universalmente las cosas comunes» (*Analíticos Segundos*, A 11, 77a26-29; aquí 'lógica' es una traducción apropiada de *dialektiké*); Frege dice que «la manera más confiable de llevar a cabo una prueba es, obviamente, seguir la lógica pura, una manera que, olvidándose de las características particulares de los objetos, depende únicamente de aquellas leyes sobre las que descansa todo conocimiento» (Frege, 1879, 5). La idea aparece muy claramente también en Tarski, 1937.

Los lógicos y filósofos de la lógica han intentado a menudo ofrecer caracterizaciones más precisas de la noción de expresión lógica que la sugerida por los vagos principios pragmáticos recién comentados. A menudo también, las caracterizaciones ofrecidas se han basado en ciertas concepciones, en mi opinión sin demasiado sustento, que atribuyen a las expresiones lógicas ciertas peculiaridades epistemológicas y modales. No entraremos en un examen crítico de estas concepciones, pero sí describiremos someramente un par de tipos de caracterizaciones de la noción de expresión lógica, que son representativos de la literatura sobre estos temas, con el objetivo principal de indicar por qué parecen deficientes.

Las caracterizaciones del primer tipo están pensadas para un ámbito de aplicación relativamente restringido, pero dentro del cual cae la cuestión acerca de si las cuantificaciones de orden superior son expresiones lógicas. En concreto, están pensadas precisamente para las expresiones de los lenguajes cuantificacionales (de orden 1 y superiores). La base de estas caracterizaciones es un concepto técnico de «noción». Este uso de la palabra se debe a Tarski. En el sentido técnico, una «noción» es un objeto de alguno de los tipos posibles en la jerarquía de objetos por medio de la que se interpretan las variables en los lenguajes cuantificacionales: los individuos son nociones, los conjuntos de individuos son nociones, los conjuntos de pares de individuos (relaciones diádicas) son nociones, los conjuntos de conjuntos de individuos son nociones, etc. (véase Tarski, 1966).

Una de las caracterizaciones del primer tipo se debe también a Tarski. Éste indicó que las funciones veritativas y las denotaciones de las cuantificaciones clásicas se pueden construir en la jerarquía de tipos como ciertas nociones que son invariantes bajo todas las permutaciones, o transformaciones 1-1, del universo de discurso. Por ejemplo, los valores de verdad «verdadero» y «falso» pueden identificarse con el universo de discurso y el conjunto vacío, respectivamente, y las funciones veritativas identificarse a su vez con funciones que tienen (tuplas de) estas clases como argumentos y valores; y las denotaciones de las cuantificaciones clásicas universales y existenciales sobre un tipo de objetos t pueden identificarse con ciertas funciones de la clase de los conjuntos de objetos de tipo t en la clase de los valores de verdad —continuando con la identificación de «verdadero» con el conjunto universal de objetos del tipo t y de «falso» con el conjunto vacío de ese tipo—. La denotación de una cuantificación universal asignará «verdadero» al conjunto de todos los objetos de tipo t, y «falso» a todos los otros subconjuntos de t; y la denotación de una cuantificación existencial asignará «verdadero» a los subconjuntos no vacíos, y «falso» al subconjunto vacío.

Una permutación de una clase induce permutaciones de todos los tipos en la jerarquía de tipos de nociones determinada por esa clase. Así, una permutación *P* de un dominio de individuos *D* induce una permutación de la clase de las relaciones *n*-ádicas de elementos de *D*, una permutación de la clase de relaciones *n*-ádicas entre relaciones de elementos de *D*, etc. Una noción u objeto *o* de un cierto tipo *t* es invariante bajo todas las permutaciones del universo de discurso si, para todas las permutaciones *P* de este universo, las permutaciones $\tilde{P}$ inducidas por *P* en la clase de las nociones de tipo *t* son todas tales que $\tilde{P}(o) = o$. Digamos que *den*(*E*, *U*) es la denotación de una expresión *E* en un universo *U*. Entonces la propiedad por medio de la que Tarski caracteriza el concepto de expresión lógica es la siguiente:

(ELT) Para todo universo *U* y para toda permutación *P* de *U*, $\tilde{P}(den(E, U)) = den(E, U)$. (Véase Tarski y Givant, 1987, 57.) (Abreviaremos esta condición diciendo '*E* es una expresión lógica tarskiana'.)

Así, *E* es una expresión lógica tarskiana cuando denota en todo universo una noción invariante bajo permutaciones del universo de discurso.

La condición (ELT) está relacionada con una condición más fuerte, que no determina necesariamente la misma extensión. Dada una biyección *B* entre dos universos *U* y *V*, usemos la notación '$\tilde{B}$' para designar la biyección inducida por *B* entre los objetos de la jerarquía de tipos generada por *U* y los objetos de la jerarquía de tipos generada por *V*. Entonces la condición más fuerte es:

(ELM) Para cualesquiera universos *U* y *V* de la misma cardinalidad y toda biyección *B* de *U* sobre *V*, $\tilde{B}(den(E, U)) = den(E, V)$. (Abreviaremos esta condición diciendo '*E* es una expresión lógica mostowskiana'[12].)

Dicho de manera intuitiva, la caracterización basada en (ELM) requiere que las expresiones lógicas denoten, en todo universo, no meramente una noción invariante bajo permutaciones de ese universo,

12. Uso esta terminología ya que Mostowski (1957) dio una definición de una cierta clase de cuantificadores generalizados de primer orden como, en esencia, los cuantificadores cuya denotación es una expresión lógica mostowskiana. Pero hay que subrayar que Mostowski no usó su condición de invariancia bajo biyecciones para dar una caracterización del concepto de expresión lógica. Esta condición ha sido usada para este propósito por otros autores, por ejemplo Sher, 1991.

sino una noción que sea la misma, salvo isomorfismo, en *todos* los universos de la misma cardinalidad. Si una expresión es una expresión lógica mostowskiana, entonces es una expresión lógica tarskiana, pues toda permutación de un universo *U* es una biyección de *U* sobre *U*[13].

No es demasiado difícil ver que todas las expresiones típicamente consideradas como lógicas en los lenguajes cuantificacionales, incluidas las cuantificaciones de orden superior, son expresiones lógicas tarskianas (y también mostowskianas). El que las cuantificaciones de orden superior sean catalogadas como expresiones lógicas por la caracterización de Tarski, naturalmente, apoya su argumento explicado al final de la sección 6. Sin embargo, este resultado no es tan cómodo como parece a primera vista, pues hay buenas razones para pensar que la caracterización de Tarski no se basa en una condición suficiente para que una expresión sea lógica en el sentido intuitivo.

Consideremos el predicado 'Hp', que podemos añadir a un lenguaje cuantificacional, con el significado intuitivo «es un heptaedro, es decir, un poliedro regular de siete caras». Este predicado denota el conjunto vacío en todos los universos reales y posibles, pues no hay ni puede haber heptaedros. Por tanto, es una expresión lógica tarskiana. (Y también una expresión lógica mostowskiana.) Pero en ninguno de los usos habituales de 'expresión lógica' compatibles con el significado de la expresión (por vago que sea) se aceptaría a 'Hp' como expresión lógica. Naturalmente, hay muchos otros ejemplos similares aunque más complicados para expresiones con denotaciones en otros tipos de la jerarquía de nociones. Estos ejemplos parecen mostrar concluyentemente que la caracterización tarskiana del concepto de expresión lógica, así como la inspirada en las ideas de Mostowski, no cumplen con el requisito de determinar un conjunto de expresiones compatible con el uso (por vago que sea) de la expresión 'expresión lógica'.

Estos conceptos técnicos de expresión lógica están muy relacionados con el concepto tarskiano de validez, no sólo por darse en términos de conceptos de la teoría de conjuntos sino, más importante aún, por el hecho de que están construidos de manera que se hace obvio que son *completos* con respecto al concepto que se busca caracte-

13. Pero no es necesariamente el caso que toda expresión lógica tarskiana sea una expresión lógica mostowskiana (contra lo afirmado en Sher, 1991, 63). Supongamos, por ejemplo, que *hubiera* una constante de predicado monádica, C, cuyo significado fuera tal que, en todo universo, C denotara o bien el conjunto vacío o el universo mismo, y que denotara el conjunto vacío en algunos universos de una cierta cardinalidad κ y el conjunto universal en otros universos de cardinalidad κ. C sería una expresión lógica tarskiana, pero no una expresión lógica mostowskiana.

rizar con ellos. En efecto, parece razonable pensar que toda expresión lógica de los lenguajes cuantificacionales habrá de ser una expresión lógica tarskiana e incluso mostowskiana. Las caracterizaciones del concepto de expresión lógica del otro tipo que comentaremos brevemente están, no por casualidad, relacionadas con el concepto de derivabilidad en un sistema formal, tanto por el uso de nociones de tipo sintáctico como por el hecho de que están construidas con la intención de que sean al menos *correctas* con respecto a las nociones intuitivas caracterizadas. Sin embargo, vamos a ver que es muy dudoso que haya una caracterización de este segundo tipo que sea al mismo tiempo correcta y completa.

Las caracterizaciones de este segundo tipo se basan en una condición cómodamente vaga que da lugar a distintos conceptos más precisos de expresión lógica según se interprete la condición de una u otra manera. La condición es la siguiente:

(KH) El significado de *E* está determinado por un conjunto de reglas de un cálculo de secuentes de Gentzen que se usan en las inferencias con *E*.

('KH' es por 'Kneale-Hacking'; la razón de este nombre se explica más abajo.) La principal fuente de vaguedad en (KH) es el significado de 'significado'. Aun suponiendo que el significado de 'estar determinado' en (KH) ha sido especificado de una manera mínimamente satisfactoria —y eso es lo que generosamente supondremos nosotros aquí que es posible hacer—, la palabra 'significado' puede interpretarse al menos de dos maneras, que han dado lugar a otras tantas caracterizaciones de la noción de expresión lógica.

Por 'significado' puede uno entender algo así como el conjunto de aspectos semánticos del uso de una expresión que es pertinente conocer para ser un usuario competente de ella, conjunto que podemos llamar 'el sentido de la expresión'. Eso parece haber entendido Kneale (1956), que es quizá quien dio la primera caracterización suficientemente clara del tipo (KH). La condición de Kneale es, pues, la siguiente:

(K) El sentido de *E* está determinado por un conjunto de reglas de un cálculo de secuentes de Gentzen que se usan en las inferencias con *E*.

Pero si uno entiende esto, entonces es difícil ver cómo evitar la objeción de que (K) no es una condición necesaria de las expresiones

lógicas intuitivas. En el caso de las cuantificaciones universales de primer orden, por ejemplo, está claro que un aspecto de su sentido es que se usan en la práctica inferencial de pasar de un examen exhaustivo de todos los casos observados de una generalización finita a la generalización misma. Pero es difícil ver cómo o en qué sentido la regla gentzeniana de introducción de las cuantificaciones universales pueda determinar que esta práctica forma parte del sentido de esas expresiones.

Una caracterización diferente basada en (KH) es la de Hacking (1979). Hacking usa el concepto de denotación para entender 'significado'. La condición de Hacking es la siguiente:

> (H) La denotación de *E* está determinada por un conjunto de reglas de un cálculo de secuentes de Gentzen que se usan en las inferencias con *E*.

Claramente puede haber expresiones con la misma denotación pero distinto sentido, así que una caracterización del concepto de expresión lógica basada en (H) no está sujeta a la objeción anterior. Hacking acompaña su caracterización basada en (H) con un resultado que afirma haber demostrado y según el cual las reglas gentzenianas de las expresiones lógicas intuitivas de los lenguajes cuantificacionales de primer orden «determinan» su denotación tarskiana (que explicamos hace unos párrafos, al hablar del concepto de expresión lógica tarskiana). También afirma que es posible demostrar que las denotaciones de las cuantificaciones de orden superior no están «determinadas» por sus reglas gentzenianas.

No es éste el lugar para examinar críticamente estas afirmaciones técnicas de Hacking, pero sí para señalar que, si es cierto que las reglas gentzenianas de las expresiones lógicas intuitivas de los lenguajes cuantificacionales de primer orden «determinan» su denotación tarskiana, entonces no cabe sino pensar que (H) no es una condición suficiente para que una expresión sea lógica en el sentido intuitivo. La razón es que es posible imaginar, o simplemente estipular, la existencia de expresiones gobernadas en parte por las mismas reglas gentzenianas que una expresión lógica intuitiva y que, sin embargo, sean intuitivamente expresiones no lógicas. Consideremos la cuantificación de primer orden 'para algún x, si no hay heptaedros, o para todo x, si hay heptaedros'[14], la cual podemos estipular como sinónima de la

14. Dicho de manera quizá más perspicua: $\underline{\exists}xF$ es verdadero si y sólo si [(no hay heptaedros y $\exists xF$) o (hay heptaedros y $\forall xF$)].

expresión '∃x', expresión que podemos tomar como un primitivo con el sentido indicado. Tiene la misma denotación tarskiana que las cuantificaciones existenciales tarskianas usuales de primer orden (en cualquier universo). A diferencia de las cuantificaciones existenciales normales, '∃x' no es intuitivamente una expresión lógica. Pero si Hacking está en lo cierto, entonces cumple la condición (H), puesto que podemos suponer o estipular que las reglas gentzenianas típicas de la cuantificación existencial gobiernan el uso de '∃x', y entonces (si la afirmación técnica de Hacking es correcta) la denotación de '∃x' está determinada por un conjunto de reglas gentzenianas que se usan en las inferencias con '∃x'. '∃x' no es una expresión lógica intuitiva, a pesar de que es una expresión lógica hackingiana.

El resultado de nuestro examen de la noción de expresión lógica y de los intentos de caracterizarla es, como adelantamos, un poco desesperanzador desde el punto de vista teórico. Pero es importante subrayar, como ya lo hicimos a propósito de la noción de consecuencia lógica, que independientemente de lo buenas o malas que sean las caracterizaciones teóricas a mano, las intuiciones existentes acerca de la noción preteórica de expresión lógica permiten muy a menudo usarla sin muchas vacilaciones. Un examen más detenido de un buen número de cuestiones relacionadas con la noción de expresión lógica se hallará en Gómez Torrente (2002).

REFERENCIAS BIBLIOGRÁFICAS

Belnap, N. D. (1962), «Tonk, Plonk and Plink»: *Analysis*, vol. 22, pp. 130-134.

Boolos, G. (1975), «On Second-Order Logic»: *Journal of Philosophy*, vol. 72, 509-527.

Etchemendy, J. (1990), *The Concept of Logical Consequence*, Harvard University Press, Cambridge, Mass.

Frege, G. (1879), *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Nebert, Halle, 1879.

García-Carpintero, M. (1993), «The Grounds for the Model-Theoretic Account of the Logical Properties»: *Notre Dame Journal of Formal Logic*, vol. 34, 107-131.

Gentzen, G. (1934), «Untersuchungen über das logische Schliessen»: *Mathematische Zeitschrift*, vol. 39, 176-210, 405-431.

Gödel, K. (1930), «Die Vollständigkeit der Axiome des logischen Funktionenkalküls»: *Monatshefte für Mathematik und Physik*, vol. 37, 349-360.

Gödel, K. (1931), «Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I»: *Monatshefte für Mathematik und Physik*, vol. 38, 173-198.

Gómez Torrente, M. (1996), «Tarski on Logical Consequence»: *Notre Dame Journal of Formal Logic*, vol. 37, 125-151.
Gómez Torrente, M. (1998), «On a Fallacy Attributed to Tarski»: *History and Philosophy of Logic*, vol. 19, 227-234.
Gómez Torrente, M. (1998-1999), «Logical Truth and Tarskian Logical Truth»: *Synthese*, vol. 117, 375-408.
Gómez Torrente, M. (2000a), «A Note on Formality and Logical Consequence»: *Journal of Philosophical Logic*, vol. 29, 529-539.
Gómez Torrente, M. (2000b), *Forma y Modalidad. Una Introducción al Concepto de Consecuencia Lógica*, Eudeba, Buenos Aires.
Gómez Torrente, M. (2002), «The Problem of Logical Constants»: *Bulletin of Symbolic Logic*, vol. 8, 1-37.
Hacking, I. (1979), «What is Logic?»: *Journal of Philosophy*, vol. 76, 285-319.
Hanson, W. H. (1997), «The Concept of Logical Consequence»: *Philosophical Review*, vol. 106, 365-409.
Hanson, W. H. (1999), «Ray on Tarski on Logical Consequence»: *Journal of Philosophical Logic*, vol. 28, 607-618.
Jané, I. (1989), «Lógica y ontología»: *Theoria*, vol. 4, 81-106.
Jané, I. (1997), «Theoremhood and Logical Consequence»: *Theoria*, vol. 12, 139-160.
Kneale, W. (1956), «The Province of Logic», en H. D. Lewis (comp.), *Contemporary British Philosophy*, 3rd Series, Allen & Unwin, Londres, 237-26.
Kreisel, G. (1967), «Informal Rigour and Completeness Proofs», en I. Lakatos (comp.), *Problems in the Philosophy of Mathematics*, North-Holland, Amsterdam, 138-171.
McGee, V. (1992), «Two Problems with Tarski's Theory of Consequence»: *Proceedings of the Aristotelian Society*, n.s., vol. 92, 273-292.
McGee, V. (1996), «Logical Operations»: *Journal of Philosophical Logic*, vol. 25, 567-580.
Mostowski, A. (1957), «On a Generalization of Quantifiers»: *Fundamenta Mathematicae*, vol. 44, 12-36.
Orayen, R. (1989), *Lógica, Significado y Ontología*, Instituto de Investigaciones Filosóficas/UNAM, México D.F.; especialmente cap. I, § 4, y cap. IV.
Prior, A. N. (1960), «The Runabout Inference-Ticket»: *Analysis*, vol. 21, 38-39.
Quine, W. V. (^{2}1986), *Philosophy of Logic*, Harvard University Press, Cambridge, Mass.
Ray, G. (1996), «Logical Consequence: a Defense of Tarski»: *Journal of Philosophical Logic*, vol. 25, 617-677.
Sagüillo, J. M. (1997), «Logical Consequence Revisited»: *Bulletin of Symbolic Logic*, vol. 3, 216-241.
Sainsbury, M. (1991), *Logical Forms*, Basil Blackwell, Oxford.
Shapiro, S. (1991), *Foundations without Foundationalism: a Case for Second-Order Logic*, Clarendon Press, Oxford.
Sher, G. (1991), *The Bounds of Logic. A Generalized Viewpoint*, MIT Press, Cambridge, Mass.

Sher, G. (1996), «Did Tarski Commit 'Tarski's Fallacy'?»: *Journal of Symbolic Logic*, vol. 61, 653-686.

Soames, S. (1999), *Understanding Truth*, Oxford University Press, Nueva York; especialmente cap. 4, apéndice.

Tarski, A. (1935), «The Concept of Truth in Formalized Languages», en Tarski (1983), 152-278.

Tarski, A. (1936), «On the Concept of Logical Consequence», en Tarski (1983), 409-420.

Tarski (1937), *Einführung in die mathematische Logik und in die Methodologie der Mathematik*, Julius Springer, Viena.

Tarski, A. (1966), «What Are Logical Notions?»: *History and Philosophy of Logic*, vol. 7, 1986, 143-154.

Tarski, A. (^{2}1983), *Logic, Semantics, Metamathematics*, Hackett, Indianapolis.

Tarski, A. y S. Givant (1987), *A Formalization of Set Theory without Variables*, American Mathematical Society, Providence, Rhode Island.

Warmbrōd, K. (1999), «Logical Constants»: *Mind*, vol. 108, 503-538.

ABSTRACCIÓN

Ignacio Angelelli

1. INTRODUCCIÓN

Durante siglos las palabras «abstracción» y «abstracto» han tenido un significado fundamental, basado en que abstraer es no prestar atención a ciertos aspectos de las cosas, mientras, naturalmente, se la sigue prestando a otros: algo es *left out* (dejado de lado), pero también algo es *retained* («retenido», Locke, 1959, III, 3, 7-8). Esto se refleja en las grandes obras de consulta general: la *Espasa* llama abstracción «a la fijación de la atención» y dice que la abstracción es «una acción mediante la cual se separa mentalmente»; la 11.ª edición de la *Britannica* describe la abstracción como *the process or result of drawing away, the process by which attention is diverted*. La misma idea queda confirmada por multitud de diccionarios y enciclopedias de filosofía. No debe alarmar que algunos diccionarios filosóficos aludan, además, a un sentido material o físico del término «abstracción», como «desplumar un ave» (*cum ab ave pennae avelluntur*, Signoriello, 1893); por lo pronto porque otros diccionarios se apresuran a decir que, por ejemplo, *la dissection d'un organe n'est pas une abstraction* (Lalande, 1947), pero sobre todo porque la referencia a la abstracción física no convierte el término en equívoco, sino que, por el contrario, muestra una extensión analógica que sólo confirma el sentido fundamental.

El sentido fundamental ha sido el único vigente durante un largo período, que puede acotarse entre Aristóteles y Peano (sección 3). Pero tras tan largo y monótono reinado, las palabras «abstracción» y «abstracto» han experimentado, en el siglo XX, una agitada vida semántica, pasando de la univocidad a la equivocidad (usos en que la idea de no prestar atención desaparece totalmente), e incluso el tér-

mino «abstracción» ha llegado a desaparecer del vocabulario filosófico (sección 4). Ante la equivocidad de un término se puede adoptar una actitud tolerante —al fin de cuentas los términos se pueden definir libremente, con tal de advertir debidamente a la audiencia o al lector—. O bien se puede señalar (según se hace aquí) que hay un sentido *genuino* (afinado y mejorado, como se propone en la sección 2) y que los otros —sentidos *espurios* o *pseudo-significados*— se deben a malentendidos o accidentes históricos. El artículo termina mencionando la conveniencia de distinguir entre *teoría* y *práctica* de la abstracción (sección 5), y formulando algunas breves conclusiones (sección 6).

2. PAUTAS PARA LA TEORÍA DE LA ABSTRACCIÓN

No basta defender el sentido genuino de «abstracción» y «abstracto» tal como de hecho se ha dado históricamente. Es preciso reforzar la noción clave —el no prestar atención— de dos maneras. Primero, tratar de que esa noción no sea blanco de las justificadas quejas antipsicologistas de, por ejemplo, Frege. Segundo, insistir en la necesidad de explicar en qué consisten los resultados de la abstracción, los supuestos entes abstractos. Esto conduce a formular los cuatro puntos o criterios siguientes:

1) *El sentido genuino del término «abstracción»*. El término «abstracción» debe entenderse estrictamente como una operación intelectual por la cual, en la consideración de las cosas, se retiene algo y se deja de lado algo.

2) *Expresión lingüística de la abstracción.* A fin de rescatar a la abstracción para la lógica, y poder someterla a un análisis riguroso, conviene estudiarla en cuanto *reflejada en los enunciados* que uno admite como verdaderos o falsos, relativos a las cosas consideradas. Durante la abstracción, algunos de los enunciados que se admitían como verdaderos se *ocultan* —lo cual *no* quiere decir que tales enunciados sean negados.

3) *Abstracción sistemática.* Si bien nada impide ocultar enunciados verdaderos de manera arbitraria, es obvio que lo normal será hacer abstracción según algún criterio determinado, lo cual se manifestará, probablemente, en la adopción de alguna *técnica abstractiva.*

4) *El producto de la abstracción*. Podríamos contentarnos diciendo que el producto de la abstracción consiste en, o en todo caso está perfectamente caracterizado por, el conjunto de enunciados que no hemos ocultado. Pero los enunciados son, en definitiva, *acerca de* ob-

jetos y lo que interesa es ver qué les pasa a objetos muy distintos cuando, bajo la luz mortecina de la abstracción, se vuelven indiscernibles. La abstracción engendra un nuevo universo de discurso, poblado por entes abstractos (*abstracta*, singular *abstractum*). La gran pregunta filosófica, que debe coronar toda teoría abstractiva, es: ¿*Qué* son los abstractos? ¿Cuál es la *naturaleza* de los abstractos? En metalenguaje: ¿Qué enunciados son verdaderos o falsos acerca de los abstractos? Al parecer, solamente dos autores: Lorenzen y Künne, han sabido formular esta pregunta de manera óptima:

> Es posible escribir una palabra en la pizarra, pero un concepto no puede ser escrito en la pizarra. Una palabra, como se suele explicar, es concreta, mientras que un concepto es abstracto. Quien sea crítico no podrá satisfacerse con esta explicación y preguntará: ¿Qué son los objetos abstractos (brevemente, los *abstracta*)? O de modo más sutil: ¿Cómo se usa el término «abstracto»? (Lorenzen, 1978, 41).

> En la filosofía analítica existe la tendencia de que a uno se lo llame platonista si supone que hay objetos abstractos. Se ha planteado la pregunta de qué condiciones se debe satisfacer para que uno sea llamado correctamente platonista, o para que se llame a una teoría platonista en el sentido indicado. En la discusión ontológica dedicada a este tema [...] la pregunta «¿qué es un objeto abstracto?» ha sido casi completamente ignorada. (Künne, 1982, 401).

Estas cuatro precisiones teóricas pueden usarse como pautas que permitan evaluar lo que unos u otros autores, a lo largo de la historia, han escrito en torno a la abstracción.

3. DE ARISTÓTELES A PEANO

La abstracción aristotélica —expresada sobre todo mediante el término αφαίρεσις— es genuina, y utiliza además una interesante técnica —la reduplicación— que sistematiza el procedimiento abstractivo, como Lear ha destacado claramente. La reduplicación consiste en añadir a un término singular la partícula «en cuanto» o equivalentes, seguida de un predicado; por ejemplo, de «Juan» se pasa a «Juan en cuanto hombre». Al hablar de un objeto x «en cuanto» poseedor de la propiedad F, *hacemos abstracción* (*filter out*, según la sugestiva expresión empleada por Lear) de todos aquellos predicados que, aunque sean verdaderos de *x*, no son verdaderos de *todos* los objetos que son F. La abstracción vía reduplicación es considerada por Lear en relación con la filosofía de la matemática en el Estagirita (también Cleary ha insistido recientemente en la especial relación que la abstracción

aristotélica tiene con la filosofía de la matemática), pero es por cierto de aplicabilidad general. Abstracción y reduplicación han ido de la mano en la historia de la filosofía (véase Angelelli, 1980).

Los investigadores (Scholz, 1935, 8; Schneider, 1970, A2; Weinberg, 1968) parecen coincidir en ver en el comentario de Boecio a las páginas iniciales de la *Isagoge* porfiriana el lugar en que comienza la historia filosófica del término latino *abstractio*. En la historia de la filosofía los *great masters of abstraction*, como dijera (en contexto peyorativo) Berkeley (1964, *Introduction*, 17) han sido los escolásticos, y esto en cada uno de los tres períodos del escolasticismo: el medieval, el moderno (paralelo a la filosofía moderna, aproximadamente entre 1500 y 1800) y el neo-escolástico.

Consideremos un ejemplo de la neo-escolástica. Santiago Ramírez, en pleno siglo XX, ha dedicado más de cien páginas, en latín, a la abstracción, dentro de su tratado *De analogia*. La detallada exposición comienza con una breve sección dedicada a la *notio abstractionis*, presentada como un movimiento de un *terminus a quo* (punto inicial) hacia un *terminus ad quem* (punto terminal), donde el punto inicial es *aliquid compositum vel coniunctum* (algo compuesto) y el terminal es la división o separación de lo que estaba unido. A continuación Ramírez despliega una enorme variedad de *divisiones* de la abstracción, cumpliendo el amenazador presagio de que *abstractio dicitur multipliciter* (la abstracción se dice de muchas maneras). Las distinciones y subdistinciones de los sentidos de la palabra *abstractio* son tan exuberantes que requieren ser impresas en una hoja mayor que las del resto del volumen, y que debe ser doblada para que no sobresalga, como suele ocurrir con los mapas —¡una verdadera cartografía de la abstracción!—. Tenemos en primer lugar la abstracción en sentido físico (*in ordine physico et corporali*, donde siempre ocurre una separación real (como cuando la nave es «abstraída» del puerto). Luego aparece la abstracción en sentido psicológico, con una distinción entre un sentido «vital» y un sentido «cognitivo»; con el segundo parece que, finalmente, hemos llegado a lo que nos interesa, aunque de nuevo se tropieza con una distinción entre *abstractio sensitiva* e *intellectiva*; la primera tiene que ver con el hecho de que un sentido no capta lo que otros sentidos captan. Promete más la abstracción intelectiva, a la que el lector se acerca con entusiasmo, aunque es recibido por una nueva distinción, ahora entre abstracción *total* y *formal*.

En la abstracción total reconocemos lo que se ha solido llamar abstracción en lógica y en filosofía, a saber, el pasaje de individuos a universales, de Pedro y Juan a hombre, y de hombre y caballo a animal, etc. La palabra «total» alude al universal visto como un todo, del

cual los individuos, o universales inferiores, son «partes». Se «abstrae de» las diferencias entre Pedro y Juan, o de las diferencias entre hombre y caballo— y así surgen los conceptos universales: hombre, animal, etc. La abstracción formal, por el contrario, no parece tan simple. Por ejemplo, en una rueda de madera se «presta atención» solamente a la forma circular, y se «hace abstracción» de la madera. No está claro si «lo que se retiene», el círculo, es algo universal o singular; si fuera lo primero, no se vería la diferencia con la abstracción total. Hay también un texto curioso de Tomás de Aquino, quien parece decir que en esta abstracción no hay algo «dejado de lado» y algo «retenido», sino que todo es retenido (*In abstractione vero, quae attenditur secundum formam et materiam, utrumque manet in intellectu; abstrahendo enim formam circuli ab aere remanet seorsum in intellectu nostro et intellectus circuli et intellectus aeris*, 1950, I 40, 3, *respondeo*).

En el marco de esta teoría abstractiva, reflejada en numerosos autores fuera de la escolástica (Descartes, Locke, Kant y tantos otros), la abstracción ha sido aplicada a dos casos notables: los *universales* y el *número*. Ambos se presentan juntos, aunque de manera muy breve, por ejemplo en Descartes, *Principia*, I, 59. El caso de los universales es muy conocido: para quienes quieren mantenerse distantes tanto del nominalismo como del platonismo, la abstracción es una conveniente estrategia. El universal *hombre* surge por retener lo común a Pedro y a María, ignorando lo que los diferencia. La historia del abordaje abstractivo del número, por el contrario, es menos conocida (véase, por ejemplo, las observaciones de Caramuel en *De abstractione arithmetica*, 1670, *Meditatio prooemialis* del capítulo *Arithmetica*); se conoce, más bien, su manifestación ulterior en autores como Cantor (en su célebre definición de cardinal: 1962, § 9) y Husserl (en su temprana *Philosophie der Arithmetik*, 1891).

¿Cómo evaluaremos a los «grandes maestros de la abstracción» en relación con los cuatro criterios que hemos establecido? Los laberintos escolásticos pueden cansarnos, pero en todos ellos hay algo muy positivo. Incluso la enumeración de significados de «abstracción» que poco nos interesan (abstracción física, abstracción sensorial...) ayuda a confirmar, por analogía, que abstraer es, ante todo y esencialmente, no prestar atención, dejar de lado algún aspecto de las cosas que consideramos —ese *leaving out* y también, naturalmente, *retaining* que dice Locke—. Cromp asegura (1980, IV, 17) que Tomás de Aquino insiste en este punto esencial: *le texte* [de Tomás Aquino] *ne cesse de rappeler que l'abstraction consiste dans une non-considération*. En otras palabras, el primer requisito — abstracción genuina— está perfectamente cumplido.

En cuanto al segundo punto, encontramos por ejemplo que Tomás de Aquino enfatiza que la abstracción se aplica a lo que está «unido en la realidad». Ahora bien, «lo que está unido en la realidad» tiene, como correlato lingüístico, un enunciado verdadero. La abstracción puede entonces describirse como un ocultar, entre los enunciados verdaderos, a algunos —sin negarlos, claro, simplemente ignorándolos—: *abstrahentium non est mendacium* (quienes abstraen no mienten, Wyser, 1947, 475).

En cuanto a técnicas abstractivas, aparte de la ya mencionada reduplicación (de origen aristotélico, pero extraordinariamente desarrollada en la escolástica), no habrá que esperar muchas en el contexto escolástico, ya que buena parte de la labor abstractiva está a cargo de las facultades *naturales* del intelecto (*agens*, *possibile*, etc.). Por cierto, no puede dejar de mencionarse el carácter defectuoso del procedimiento abstractivo que se ha solido emplear para pasar de conjuntos concretos a sus números. Se tiende a describir ese procedimiento como un «no atender» a la «naturaleza» de los elementos del conjunto concreto. Esto se ha reflejado en autores de gran importancia, fuera del ámbito escolástico, como Descartes, Cantor, Husserl (véanse textos citados más arriba). Ahora bien, según ha señalado Frege (1884, § 34), al «desatender» las diferencias entre los elementos, lo que «queda» no parece ser el ansiado número sino simplemente el concepto universal común a esos elementos; según su famoso ejemplo, si se parte de dos gatos, uno blanco y otro negro, y se ignoran las diferencias y peculiaridades de cada gato, se llegará a un cierto concepto de gato, que reúne lo que es común a ambos gatos iniciales, más bien que a algo que merezca ser llamado «2». Frege no tuvo interés en reparar el defectuoso mecanismo de la abstracción tradicionalmente aplicada en el caso del número, y prefirió desechar la abstracción. Ha sido Lorenzen (1955) quien por vez primera vez —en el marco de la técnica iniciada por Peano, y coincidiendo con sugerencias de Weyl (1928, § 2, ejemplo 2)— ha reparado de manera sistemática el antiguo desacierto.

Finalmente, en cuanto al cuarto punto, la naturaleza de los *abstracta*, se puede vislumbrar qué contestaría un escolástico, o un autor clásico —entre Aristóteles y Peano— a la pregunta ¿qué es el abstracto? En el caso de los universales, diría que se trata de una colección de propiedades; hombre, por ejemplo, es el conjunto {animal, racional}. Ésta es la famosa «naturaleza considerada en sí misma». La defectuosa teoría aristotélica de la predicación, heredada por los escolásticos y vigente hasta la revolución fregeana (véase Angelelli, 1975), ha sido un impedimento para permitir el fácil discurso acerca de es-

tos abstractos, y por consiguiente para llegar a cierta claridad en cuanto a qué es lo que puede decirse acerca de ellos, en este caso acerca de la «naturaleza considerada en sí misma». En el caso de los números, la respuesta que tiende a perfilarse en la tradición filosófica —incluyendo a Cantor y al temprano Husserl— es que el abstracto, o sea el resultado de la abstracción, es como un conjunto de palidísimas unidades. Aquí la objeción fregeana, mencionada más arriba, es fatal: ignorando las diferencias entre los elementos del conjunto concreto inicialmente dado, no quedan unidades «abstractas», ni siquiera dos palidísimos gatos, sino un concepto universal, que reúne lo que es común a los elementos concretos iniciales.

Cuando ya no parece haber sitio para novedades, hacia finales del siglo XIX acontece algo inusitado en la teoría de la abstracción. El matemático Peano esboza una técnica abstractiva inédita. Consideremos el universo de discurso de los quebrados: {1/2, 3/8, 5/7, 3/6...} y consideremos, para los objetos que pueblan este universo de discurso, la relación de tener iguales productos cruzados. En esta relación están 1/2 y 3/6, pero no 1/2 y 5/7. Propongámonos ahora llevar a cabo una abstracción de manera sistemática, como sigue. Al hablar acerca de cualquier objeto de nuestro universo, nos impondremos silencio toda vez que lo que estemos a punto de decir no sea también verdadero acerca de cualquier otro objeto que se encuentre con el dado en la relación que hemos descrito. Así, por ejemplo, aunque el enunciado «1/2 es irreductible» es perfectamente verdadero, la ascesis abstractiva que nos hemos impuesto nos prohíbe decirlo; no se trata de negarlo, sino simplemente de ocultarlo, «hacer abstracción» de él (Peano, 1958, II, 172).

Peano satisface, de manera excelente, las cuatro pautas formuladas en la sección 2 de este artículo. No sólo hay en Peano hay una novedosa técnica que permite ejercer el antiguo «no prestar atención» de manera que obviamente escapa a las objeciones antipsicologistas (por ejemplo las fregeanas), sino que hay también un intento al menos de decir qué es un ente abstracto, claramente reconocido como *nouvel objet* (el *abstractum* es para Peano un conjunto de propiedades: *ibid.*, 169, también 410), y por añadidura una clara voluntad de enlace con la tradición filosófica clásica. Como bien lo pusiera de manifiesto Scholz (1935, 2) citando la siguiente afirmación de Peano: «La parola "astrazione" di uso internazionale fu introdotta in filosofia da Boezio, per tradurre il greco ἀφαίρεσις» (Peano, 1958, 409). Así, no es difícil entusiasmarse y ver a Peano como autor de una teoría abstractiva «moderna», mejor que la «clásica» (Angelelli, 1981).

4. ABSTRACCIÓN EN EL SIGLO XX: EXILIO, PSEUDO-USOS, MINORITARIA SUPERVIVENCIA DE LA ABSTRACCIÓN GENUINA

Si bien la gran mayoría de las obras de consulta, generales o especialmente filosóficas, que ocupan las bibliotecas actuales explican la abstracción en el sentido narrado en la sección anterior (no prestar atención), hay algunos hechos curiosos que pueden intrigar a quien se esté iniciando en el estudio del término.

El término *abstracción* ha desaparecido de la Enciclopedia Británica en su 15.ª edición, y tampoco figura en la enciclopedia filosófica Routledge, la más reciente y la máxima producida hasta ahora en el idioma inglés, que sólo ofrece el adjetivo *abstracto* (lo mismo ocurre en la enciclopedia electrónica Stanford, pero en este caso no se puede emitir un juicio definitivo, tratándose de una obra «dinámica», en crecimiento constante).

En el diccionario filosófico Runes (1961) hay un artículo («Abstraction»), de menos de una página, compuesto de tres notas firmadas por autores diferentes. La primera y la tercera nota repiten lo consabido: abstraer es separar, idealmente, un aspecto de una cosa, concentrando la atención en ese aspecto, y llegando así a ver lo que cosas diferentes tienen en común. La segunda nota, firmada por Alonzo Church, explica dos procedimientos a los que el célebre lógico llama abstracción. El primero es el pasaje de un dominio de objetos provisto de una relación de equivalencia a otro dominio sobre la naturaleza de cuyos elementos Church no nos dice absolutamente nada, pero que caracteriza estructuralmente así: a dos objetos del dominio inicial se le asigna un objeto del dominio nuevo si y sólo si los dos objetos están en la relación de equivalencia en cuestión. El segundo proceso es puramente simbólico: dada una expresión con variables libres, se puede «cerrar» a las variables, creando un nombre de un ente nuevo. Es obvio que ninguno de los dos procedimientos llamados «abstracción» por Church exhibe el menor indicio de «no atender a».

Nada de esto es casual. Los hechos indicados nos ponen en la pista de lo que le ha pasado a la abstracción en el siglo XX, dentro del marco de lo puede describirse, globalmente, como *mainstream* de la filosofía analítica, lógica, y estudios de fundamentos de la matemática. La abstracción ha sido *exilada* de ese *mainstream*, y el lugar vacante ha sido ocupado por *pseudo-usos* del término —*usurpadores*, si queremos completar la metáfora del exilio—. Dos causas, al menos, han contribuido, conjuntamente, a que se produjeran ese doble fenómeno: 1) Frege, muy justificadamente, vio los defectos —pero la-

mentablemente *sólo* los defectos— de la teoría abstractiva, aplicada al caso especial que le interesaba: el número, y en vez de corregir la teoría abstractiva prefirió eliminarla (véase Angelelli, 1984). 2) Peano, pese a haber acumulado títulos como para ser clasificado entre los mejores teóricos de la abstracción en la historia de la filosofía, no supo llevar a buen término su empresa, y acabó generando confusión entre sus seguidores y autores posteriores.

Un ejemplo notable del «exilio» se encuentra en el desarrollo personal de Russell, quien luego de hablar reiteradamente de un *principle of abstraction* pasó a hablar de un principio que permite *dispense with abstraction* (1956, 326), parecidamente a lo que se observa en la evolución de la *Britannica*.

En cuanto a los usos espurios, pueden distinguirse al menos tres pseudo-usos de «abstracción» y uno del adjetivo «abstracto» (véase Angelelli, 1991). Empezando por este último, se suele llamar «abstracto» a algo que está fuera del espacio y del tiempo. Se advierte el contraste entre este pseudo-uso de «abstracto» y el genuino, observando que, para un amigo de las formas platónicas o para un creyente, las formas o Dios son abstractos en el sentido espurio, pero de ninguna manera en el sentido auténtico: ni el platónico ni el creyente aceptan que las formas o Dios resulten de no prestar atención. Dentro de los tres pseudo-usos de «abstracción» hay uno «mayor» y dos «menores». De los dos menores ya hemos encontrado uno en la nota de Church (1961) para el diccionario Runes; otro es el que H. Scholz (1935, 15) ha denominado *logistische Abstraktionstechnik*, y que consiste en pasar de un enunciado a un predicado, reemplazando algún término singular del enunciado mediante una variable; por ejemplo de «Pedro es hombre» a «*x* es hombre». Es obvio que en ninguno de estos dos casos «menores» es preciso ejercer abstracción alguna; se trata de operaciones con símbolos, donde a lo sumo puede decirse, con benevolencia, que la expresión resultante designa entes abstractos, desplazando entonces la cuestión a ¿por qué se llama abstractos a esos entes?, ¿por qué son productos de algún tipo de abstracción genuina o por qué no pueden «tocarse», estando fuera del espacio y del tiempo? El pseudo-uso «mayor» de «abstracción», paradójicamente, se origina, en cierto modo, en Peano, y requiere comentarios especiales.

Retomemos la consideración de la técnica abstractiva peaniana, y sigamos considerando el mismo ejemplo: un universo de discurso constituido por los quebrados, o parejas de números enteros positivos: {1/2, 3/8, 5/7, 3/6...}, donde por ejemplo 1/2 y 3/6 son muy diferentes, pero donde súbitamente decidimos limitar nuestros enunciados a aquellos que son invariantes con respecto a la relación de

tener iguales productos cruzados. Este semi-mutismo conduce a que, por ejemplo, 1/2 y 3/6 se vuelvan *indiscernibles*. Evidentemente, el universo de discurso ha cambiado, ahora hablamos de *otra* cosa. Antes de averiguar la naturaleza de los nuevos entes, bauticémoslos. Por ejemplo, prefijemos una tilde «~» a los términos singulares que designaban los objetos de nuestro universo inicial; así «~1/2» denota ahora a ese ente que es indiscernible —¡mientras dure nuestra abstracción!— de ~3/6. Es obviamente verdadero que, *bajo la abstracción*, ~1/2 = ~3/6, y de modo más general, que ~a/b = ~c/d si y solamente si a/b y c/d están en la relación elegida (o sea: a.d = b.c). Generalicemos aún más, y consideremos *cualquier* universo de discurso {a, b, c...}, donde a, b, c ya no tienen que ser quebrados o números sino que son objetos absolutamente cualesquiera, para los cuales hemos definido una relación x ~ y de *equivalencia* cualquiera (de que tener iguales productos cruzados es un mero ejemplo). Impuesto el mismo ascetismo abstractivo, llegaremos a nuevos términos singulares «~a», «~b», y aceptaremos como obviamente verdadero —mientras dura la abstracción— al bicondicional ~a = ~b ↔ a ~b. Ahora bien, el enorme defecto, si no en el pensamiento (que no podemos conocer) en todo caso en los escritos de Peano, consiste en dar la impresión de que la equivalencia ~a = ~b ↔ a ~b no es para él un enunciado verdadero que surge *mientras se ejerce* la teoría abstractiva, sino una mera estipulación o definición que uno puede hacer toda vez que disponga de una relación de equivalencia, donde el *definiens* es a ~b y el *definiendum* es ~a = ~b. A esta definición Peano la llama *definizione per astrazione* (véase Peano, 1958, II, 413). Es curioso y lamentable que teniendo todos los elementos para una excelente teoría abstractiva, incluso una gran conciencia histórica, Peano, sea por presión del entorno, sea por confusión o por debilidad filosófica (manifestada, por ejemplo, en la excesiva tolerancia de la *Conclusione* del ensayo *Le definizioni per astrazione* (Peano, 1958, II, 409), no supo ver al bicondicional «~a = ~b ↔ a ~b» como verdadero sin más, y dejó que circulara como una estipulación aún inconclusa, donde falta determinar qué es ~a y qué es ~b. El resultado fue catastrófico, sobre todo por una rara coincidencia que vino a complicar, desde fuera, la historia de la abstracción, y que pasamos a narrar.

Frege había ya iniciado (1884) un nuevo método, aplicándolo por primera vez a su tema central: el número. Este método tiene peligrosos parecidos con la abstracción peaniana, si bien no es abstracción ni pretende serlo. Frege quiere saber qué es el número, y encuentra que todo número se manifiesta como asociado a un concepto, de modo que sin restricción alguna de la generalidad, la pregunta ¿qué es el nu-

mero? puede transformarse en «¿qué es el número de F?» (F = un concepto). O sea, que la necesidad inicial de esclarecer la naturaleza del número se convierte en la necesidad de esclarecer el significado de los términos singulares de la forma: «el número del concepto F». Inseguro al abordar directamente tamaña cuestión, Frege procede, cautelosamente, en dos etapas. Primero fija una condición para que dos números, cada uno naturalmente asociado con un concepto, digamos N'F y N'G, sean idénticos, posponiendo la cuestión de la naturaleza de N'F o N'G. La condición es expresada así: N'F = N'G si y solamente si se da, entre F y G, la relación de que es posible establecer una biyección entre los objetos que caen bajo cada uno de los dos conceptos. Esta relación es de equivalencia. Así vemos que Frege empieza por un bicondicional de forma «~a = ~b ↔ a ~b». Claro que no se sabe nada acerca de lo denotado por los términos singulares «~a», «~b». En la segunda etapa de su método Frege pasa a asignar tales denotaciones, lo cual es una forma metalingüística de decir: pasa a decir qué es el número. Como se sabe, Frege elige como denotación las clases de equivalencia. Ahora bien, en ningún momento se ve a Frege ejerciendo abstracción alguna. Tampoco se percibe necesidad alguna en su elección de las clases de equivalencia. Esta arbitrariedad en la elección se vuelve hirientemente explícita en la generalización del método fregeano que, décadas más tarde, lleva a cabo Carnap (1960). Carnap describe la segunda etapa como un *looking around* (mirar en torno), para encontrar objetos que sean «apropiados» para desempeñar el papel de ser denotados por los términos singulares «~a», «~b», etc. Ser apropiado significa meramente que la elección sea compatible con el bicondicional establecido en la primera etapa. He llamado al método de Frege, en la versión carnapiana, «método de mirar en torno» (Angelelli, 1979) o, aprovechando la etimología latina, de circunspección.

Tanto en Peano como en Frege tenemos el bicondicional «~a = ~b ↔ a ~b» con su relación de equivalencia. Peano había llegado a este bicondicional *sabiendo* qué es ~a y ~b, y así el condicional es un enunciado verdadero relativo al proceso abstractivo; podemos decir que el bicondicional es *punto de llegada*. Frege no llega al bicondicional sino que empieza con él: es su *punto de partida*, sin saber aún qué es ~a o ~b. El bicondicional en realidad es un esquema regulativo, que dice que cualquiera que sea la naturaleza de ~a o de ~b, esa naturaleza tendrá que ajustarse a lo exigido por el bicondicional.

El fenómeno extraño fue que, si bien la abstracción genuina se le escapó de las manos a Peano, su frase «definizione per astrazione» continuó circulando (véase, por ejemplo, Natucci), para terminar de-

signando al método circunspectivo de Frege y Carnap, engendrando así el máximo pseudo-uso del término «abstracción», que hemos encontrado en Alonzo Church, y que sigue exhibiendo una poderosa vigencia en obras recientes como Dummett (1991; véase Angelelli, 1994). Una novedosa variante (dentro del llamado «neologicismo fregeano») es llamar abstracción simplemente a la primera parte del método («principio de Hume», véase Hale, 2000). El primero en denunciar esta confusión de los dos métodos —la circunspección originada en Frege y la abstracción originada en Peano— parece haber sido Lorenzen (1955, 101).

Es meritorio intentar, como Ladrière, rescatar las posibilidades abstractivas latentes en el método de mirar en torno. Y es correcto, porque detrás del énfasis carnapiano en que no interesa cuál sea el objeto elegido como denotación de «~a» sino simplemente interesa que el objeto elegido sea compatible con el bicondicional inicial, se descubre claramente una oscura voluntad de hacer abstracción. Pero esta voluntad debería entonces *explicitarse* y el método de mirar en torno debería ser sustituido por el método abstractivo que Peano supo diseñar pero no imponer a quienes le siguieron.

Los pseudo-usos indicados han dominado la escena lógico-analítica del siglo XX. Dentro de esa escena ha habido una excepción. La genuina (aunque históricamente abortada) preocupación abstractiva de Peano reapareció en H. Weyl y luego en P. Lorenzen. Éste, sobre todo, ha mantenido encendida la antorcha de la abstracción genuina en su «escuela de Erlangen». Su nominalismo, sin embargo, no propició que se ocupara demasiado de la cuestión, filosóficamente fundamental, de la naturaleza de los *abstracta*, pese a la forma óptima en que supo plantear la pregunta (texto citado en la sección 1; también en relación con el texto de Künne (*ibid.*), se echa de menos una respuesta satisfactoria). La abstracción genuina pregonada por Lorenzen a través de todos sus trabajos y su actividad docente ha tenido no sólo adherentes sino fuertes críticos, entre los que hay que mencionar en primer lugar a Geo Siegwart.

Fuera de este marco de lógica, filosofía analítica y fundamentos de la matemática, la abstracción genuina continuó por cierto su vida milenaria en escuelas tradicionales, como la neo-escolástica y los variados descendientes de Hegel, idealistas o dialéctico-materialistas.

No es siempre fácil clasificar como espurio o genuino lo que se escribe sobre abstracción. Quede como ejercicio para el lector interesado el determinar a cuál de los dos sectores pertenece la obra de K. Fine (1998), que manifiesta un gran entusiasmo abstractivo (*Abstraction, as Cantor and Dedekind conceive it, is ontologically innovative—*

it leads to objects that are genuinely new, 601) y una clara voluntad de obtener un *clear and definite understanding of what abstraction is* (608).

5. ABSTRACCIÓN APLICADA

Finalmente, es preciso distinguir entre *teoría* de la abstracción y *práctica* o *aplicación* de la abstracción (se entiende, *genuina*). En lo anterior hemos atendido solamente a la teoría. Es importante observar que pese a la proliferación de sentidos espurios, la abstracción genuina ha seguido manifestando gran vitalidad a nivel práctico o aplicado, de lo cual hay numerosos ejemplos distribuidos en los más variados temas: computación, arte, psicología y otros. A veces, quienes *aplican* la abstracción ni se molestan en nombrarla. Un caso notabilísimo de esto —abstracción no nombrada pero *practicada*, y perfectamente *genuina*— se encuentra en la pequeña y clásica introducción a la teoría de conjuntos de Kamke (véanse las definiciones de conjunto, cardinal, ordinal). También hay quienes la nombran con el fin de rechazarla pero no pueden despojarse enteramente de ella, y acaban aplicándola sin, al parecer, advertirlo (Frege, 1903, § 146).

6. CONCLUSIONES

A manera de conclusión, se pueden formular las tres observaciones siguientes:

1) Si bien la imposición de nombres es arbitraria, y en este sentido no cabe rechazar los usos espurios de «abstracción» y «abstracto», el hecho de que tales usos proceden de accidentes o confusiones históricas sugiere que sean desechados (introduciendo otros términos para expresarlos, si es preciso). Así, lo razonable con respecto a «abstracción» y «abstracto» es o bien usar estos términos de manera genuina, o bien no usarlos. En este sentido es preferible que una obra de consulta (general o específicamente filosófica) elimine, como hemos visto en dos casos, el término «abstracción», a que propicie su pseudo-uso.

2) Quienes, a nivel de *teoría* abstractiva, optan por borrar la abstracción de la terminología filosófica, igual que quienes se complacen en adoptar usos espurios, eluden el estudio de la abstracción genuina que, de manera innegable y pese a los teóricos, se sigue manifestando a nivel *aplicado*, como fenómeno que requiere explicación.

3) Para quienes aceptan a la abstracción (genuina) hay un reto principal: entender exactamente *qué* es un ente abstracto, *cuál* es la naturaleza de los entes abstractos, o, de manera lingüística, *cuáles* enunciados son verdaderos o falsos acerca de un presunto ente abstracto.

REFERENCIAS BIBLIOGRÁFICAS

Angelelli, I. (1975), *Freges Ort in der Begriffsgeschichte*, en Ch. Thiel (ed.), *Frege und die moderne Grundlagenforschung*, Hain, Meisenheim am Glan, 9-22.

Angelelli, I. (1979), «Abstraction, looking-around and semantics»: *Studia Leibnitiana*, Sonderheft, 8, 108-123.

Angelelli, I. (1980), «Abstracción y reduplicación»: *Revista Latinoamericana de Filosofía*, 6, 255-256.

Angelelli, I. (1981), «Abstracción moderna y tradicional»: *Anuario Filosófico* (Navarra), XIV, 2, 9-21.

Angelelli, I. (1984), «Frege and abstraction»: *Philosophia Naturalis*, 21, 453-47.

Angelelli, I. (1991), «La abstracción en la filosofía contemporánea»: *El hombre: inmanencia y transcendencia*, Servicio de Publicaciones de la Universidad de Navarra, I, 167-180.

Angelelli, I. (1994), «Abstraction and number in Michael Dummett's Frege. Philosophy of mathematics»: *Modern Logic*, 4, 308-318.

Berkeley, G. (1964), *A treatise concerning the principles of human knowledge. The works of George Berkeley Bishop of Cloyne*, Nelson, London.

Británica [Enciclopedia], artículo «Abstraction», en *The Encyclopaedia Britannica*, 11.th ed., Cambridge, 1910.

Británica [Enciclopedia], artículo «Abstraction», en *The Encyclopaedia Britannica*, 15.ª ed., Cambridge, 1974, 1998.

Cantor, G, (1962), «Beiträge zur Begründung der transfiniten Mengenlehre», en Zermelo, E. (ed.), *Georg Cantor Gesammelte Abhandlungen*, Olms, Hildesheim.

Caramuel, I. (1670), *Mathesis biceps*, Campaniae.

Carnap, R. (1960), *Meaning and necessity*, The University of Chicago Press, Chicago.

Church, A. (1961), parte del artículo «Abstraction», en Runes (1961).

Cleary, J. J. (1982), *Aristotle's theory of abstraction: a problem about the mode of being of mathematical objects*, Doctoral Diss. Boston University, University Microfilms International, Ann Arbor.

Cromp, G. (1980), *L'abstraction de l'intellect agent dans la Somme de Théologie de Thomas d'Aquin*, 4 vols., Pontifical Institute of Medieval Studies, Toronto.

Descartes. R. (1957), *Principia Philosophiae*, en Ch. Adam y P. Tannéry (eds.), *Oeuvres de Descartes*, Vrin, Paris.

Dummett, M. (1991), *Frege. Philosophy of mathematics,* Harvard University Press, Cambridge, Mass.

Espasa [Enciclopedia], artículo «Abstracción», en *Enciclopedia Universal Ilustrada Europeo Americana*, Espasa editores, Barcelona, sin fecha.
Espasa [Enciclopedia], artículo «Abstracción», en *Enciclopedia Universal Ilustrada Europeo Americana,* Apéndice A-Z, Espasa-Calpe, Madrid, 1996.
Fine, K. (1998), «Cantorian abstraction: a reconstruction and defense»: *The journal of philosophy*, 95, 12, 599-634.
Frege, G. (1884), *Die Grundlagen der Arithmetik*, Koebner, Breslau.
Frege, G. (1903), *Die Grundgesetze der Arithmetik*, vol. 2, Pohle, Jena.
Hale, B. (2000), «Reals by abstraction»: *Philosophia Mathematica*, 3, vol. 8, 100-123.
Husserl, E. (1891), *Philosophie der Arithmetik*, Halle-Saale, Pfeffer.
Kamke, E. ([4]1962), *Mengenlehre,* W. de Gruyter, Berlin.
Künne, W. (1982), *Criteria of abstractness,* en Smith, B. *et al.* (eds.), *Parts and moments*, Philosophia, Munich, 401-437.
Ladrière, J. (1989-1998), «Abstraction», en *Encyclopédie philosophique universelle,* Presses Universitaires de France, Paris.
Lalande, A. ([5]1947), *Vocabulaire technique et critique de la philosophie*, Presses Universitaires de France, Paris.
Lear, J. (1982), «Aristotle's philosophy of mathematics»: *The philosophical review*, 91, 161-192.
Locke, J. (1959), *An essay concerning human understanding*, Dover, New York.
Lorenzen, P. (1955), *Einführung in die operative Logik und Mathematik*, Springer, Berlin.
Lorenzen, P. (1978), *Theorie der technischen und politischen Vernunft*, Reclam, Stuttgart.
Natucci, A. (1923), *Il concetto di numero e le sue estensioni*, Bocca, Torino.
Peano, G. (1958), *Opere Scelte*, 3 vols, Edizioni Cremonese, Roma.
Ramírez, S. (1971), *De analogia*, CSIC, Madrid.
Routledge Encyclopedia of Philosophy, Routledge, London, 1998.
Runes, D. D. (1961), *Dictionary of philosophy*, Littlefield, Adams & Co., New Jersey.
Russell, B. (1956), «Logical atomism», en R. Ch. Marsh (ed.), *Logic and Knowledge*, Allen & Unwin, London, 321- 343.
Schneider, H. J. (1970), *Historische und systematische Untersuchungen zur Abstraktion*, Inaugural-Dissertation, Universidad de Erlangen-Nürnberg.
Scholz, H. y Schweitzer, H. (1935), *Die sogenannten Definitionen durch Abstraktion. Eine Theorie der Definitionen durch Bildung von Gleichheitsverwandtschaften*, Leipzig, 108 pp. (Según se descubre en el prefacio, este trabajo se compone de dos partes: la primera, de Scholz, tiene como título completo *Die sogenannten Definitionen durch Abstraktion und ihre Vorläufer in der klassischen Logik.*)
Siegwart. G. (1993), «Zur Inkonsistenz der konstruktivistischen Abstraktionslehre»: *Zeitschrift für philosophische Forschung*, 47, 2, 246-260.
Signoriello, N. (1893), *Lexicon peripateticum philosophico-theologicum*, Napoli.
Stanford Encyclopaedia of Philosophy (2000), http://plato.stanford. edu.

Tomás de Aquino (1950), *Summa Theologiae,* Marietti, Roma.
Weinberg, J. (1968), «Abstraction in the formation of concepts», en P. P. Wiener (ed.), *Dictionary of the history of ideas,* New York, I, 1-9.
Weyl, H. (1928), *Philosophie der Mathematik und Naturwissenschaft,* Oldenbourg, München.
Wyser, P. (1947), *Die wissenschaftstheoretischen Quaest. V u. VI in Boethium de Trinitate des hl. Thomas von Aquin.* En *Divus Thomas,* 25, 437-485, 26, 74-98.

ENTIDADES ABSTRACTAS*

Eduardo Alejandro Barrio

Es usual sostener que un argumento es válido si y sólo si todo modelo de sus premisas es un modelo de su conclusión. También lo es decir que hay proposiciones verdaderas en todos los mundos posibles o que el número de fórmulas bien formadas de la lógica proposicional es infinito. A veces discutimos sobre si la verdad lógica es una propiedad de ciertas oraciones o sobre si existe el conjunto de todos los conjuntos. La comprensión literal de tales afirmaciones parece comprometernos con modelos, proposiciones, mundos posibles, números, propiedades y conjuntos. Por supuesto, no resulta difícil pensar en afirmaciones que parezcan remitir a conceptos, funciones u otras tantas entidades semejantes. El objetivo de este trabajo es ofrecer una visión general de las discusiones más importantes involucradas en nuestro uso de afirmaciones pertenecientes a la lógica que parecen suponer un compromiso con entidades abstractas. Intentaré ofrecer una descripción de los distintos enfoques ontológicos que tradicionalmente se han elaborado alrededor de la aceptación de este tipo de entidades.

La estructura del artículo es la siguiente. Comenzaré mostrando que la lógica parece hablar de entidades cuya naturaleza excede lo que intuitivamente consideramos como entidades concretas. ¿Qué es lo que hace que sean verdaderas nuestras afirmaciones lógicas que parecen comprometernos con entidades que presuntamente están más allá de lo concreto? A continuación ofreceré diversos criterios, que nos permitirían alejarnos del plano intuitivo, para considerar abstrac-

* Doy las gracias a Gonzalo Rodríguez-Pereyra, a Marco Ruffino y a Alberto Moretti por su colaboración.

ta una entidad. La captación perceptual, la locación espacio-temporal, la capacidad de movimiento, así como la dependencia ontológica de otras entidades y su carácter presuntamente secundario en la jerarquía de lo que hay, serán algunos de los aspectos que consideraremos. Mostraré que ninguno de estos criterios resulta completamente satisfactorio. Finalmente, analizaré uno de los principales argumentos para respaldar nuestra creencia en entidades abstractas. Prestaré atención a la tesis de la indispensabilidad de las entidades abstractas.

1. ¿DE QUÉ HABLA LA LÓGICA?

Un tratamiento completo de las nociones centrales de la lógica parece requerir el recurso a entidades con una naturaleza algo misteriosa. La lógica es principalmente una teoría acerca de los razonamientos y las leyes lógicas. Al elaborar un sistema lógico, nos proponemos encontrar fundamentalmente lo que hace aceptable aquel tipo de procedimiento que proporciona razonamientos; y al mismo tiempo, tratamos de presentar un análisis de aquellos elementos que cumplen un papel central en el establecimiento de las mencionadas leyes. Un elemento clave en esa búsqueda es el reconocimiento de que, en parte, esos aspectos tienen que ver con elementos formales o estructurales. Y en este sentido, lo que es claro es que no parece posible describir en detalle formas lógicas o estructuras sin aclarar cuáles son los componentes que la integran. En las presentaciones usuales de la disciplina, tanto leyes como argumentos involucran fórmulas. En el caso de estos últimos, ellos son secuencias de fórmulas. En ambos casos, esas fórmulas contienen como componentes elementos que usualmente son denominados *constantes* y *variables lógicas*[1]. La pregunta que surge fácilmente es la siguiente: ¿Qué tipo de entidad es una fórmula?[2]. Este problema se relaciona con el de los portadores de verdad y usualmente en esas discusiones se recurre a las oraciones o a las proposiciones para darle una respuesta[3].

1. Por supuesto, también hay elementos auxiliares como los paréntesis que sirven como signos de puntuación.

2. Llamaré *fórmula* a cualquier cadena de expresiones de un lenguaje artificial que responda a las reglas de la gramática de ese lenguaje.

3. Llamaré *oración* a cualquier cadena de expresiones del lenguaje natural gramaticalmente correcta y completa. Llamaré *proposición* a lo que es común a un conjunto de oraciones declarativas sinónimas. También se habla, aunque con menos frecuencia, de *afirmaciones, creencias, enunciados* y *juicios*. Véase, por ejemplo, R. Orayen, 1989.

Antes de evaluar ambas alternativas, conviene advertir lo siguiente: en un sistema de lógica se supone que los lenguajes contienen infinitas fórmulas que cumplen el papel de ser portadores de verdad. Esta suposición obliga a establecer las condiciones de verdad de tales formulas pertenecientes al sistema de manera recursiva. Frente a estas consideraciones, Orayen argumenta correctamente que sean cuales fueren esas entidades, estamos obligados a «considerar estos portadores de verdad como tipos y no como casos, ya que no hay garantía alguna de que vaya a existir alguna vez un número infinito de inscripciones (casos) de un lenguaje dado» (Orayen, 1989, 53).

Una vez aclarado este punto, conviene recordar que toda fórmula caso está localizada en determinadas coordenadas espacio-temporales. Ella es una entidad física: tal vez unas marcas sobre un papel. Por lo anterior, dos inscripciones de la misma fórmula no son la misma fórmula caso. Son dos fórmulas caso distintas. Por supuesto, no podemos negar que ellas tengan parecido, pero estrictamente no son el mismo objeto físico. Ellas son entidades particulares, no son universales. Una fórmula tipo estrictamente hablando no está en ninguna parte del espacio-tiempo. Puede tener ejemplos o realizaciones. Esos ejemplos son fórmulas caso. Pero de ninguna manera es necesario que una fórmula tipo se ejemplifique. Es totalmente posible pensar que hay fórmulas tipo que carecen de ejemplos a lo largo de todo el espacio-tiempo.

Volvamos entonces a la cuestión que nos preocupa. No es fácil resolver la cuestión de qué tipo de entidad son las fórmulas. ¿Qué son las fórmulas tipo? Normalmente un teorema como 'p v -p' se entiende que vale para todas las instancias de 'p'. Podemos ver a 'p' como una genuina variable. En este caso, lo que se hace es considerar a las presentaciones habituales de la lógica proposicional como una versión abreviada de la lógica de cuantificacional[4]. En esta alternativa, surge inevitablemente la cuestión de la interpretación de los cuantificadores. Si se adopta la interpretación objetual, se nos presenta la cuestión de qué clase de objetos recorren los cuantificadores. ¿Tenemos oraciones o proposiciones en el dominio de interpretación? Reparos ontológicos pueden inclinarnos a pensar, en principio, que los objetos que substituyen a las variables son oraciones y no proposiciones[5]. ¿Pueden las

4. Quine propone un camino alternativo: ver a 'p', etc., no como legítimas variables capaces de ser ligadas, sino, en lugar de ello, como *letras esquemáticas*. Una fórmula de un sistema lógico, tal como 'p v -p', se la entiende no como una oración o proposición, sino como un esquema tal que todas las oraciones reales de la forma descripta se cumplan o valgan. Pero ésta no es la opción más seguida. Véase W. Quine, 1953, 109.

5. Para una defensa de las oraciones véase A. Moretti, 1984.

oraciones ser legítimos portadores de verdad? Distintos autores han afirmado que no[6]. Uno de los argumentos que se presenta es que si las oraciones tuvieran valores veritativos, algunas serían a veces verdaderas y otras falsas. La oración «llueve» altera sus valores veritativos en distintos instantes temporales. Y tal cosa resultaría inaceptable. Otro es que no todas las oraciones pueden tener valores veritativos. Preguntas, órdenes y hasta quizás oraciones que expresan conceptos morales, estéticos o políticos podrían carecer de condiciones de verdad[7]. Sin embargo, ninguno de los dos argumentos parece concluyente: el primero podría responderse relativizando la atribución de verdad al contexto de emisión de la oración[8]. Claro que por las razones ofrecidas por Orayen en cuanto al número potencial de portadores dentro de un sistema de lógica, el compromiso tendrá que ser con tipos[9]. El segundo argumento puede responderse diciendo que no hace falta comprometerse con la idea de que *todas* las oraciones tienen valor veritativo. Si algunas lo tienen, es suficiente para la propuesta. Aun así quizás podamos pensar que es mejor abandonar los reparos ontológicos y pensar que esos objetos puedan llegar a ser proposiciones.

Es usual decir que una oración expresa la proposición que constituye su significado. Así la oración española «la nieve es blanca» expresa la misma proposición que la oración inglesa «snow is white». Una misma oración puede expresar proposiciones diferentes en diferentes contextos. Por ejemplo, la oración «Está nevando» emitida en Buenos Aires expresa casi seguro una proposición falsa, pero emitida en la Antártida casi seguro que expresa una proposición verdadera.

Ahora bien, una razón para dudar acerca de las proposiciones es que ellas parecen tener dudosas condiciones de identidad. En el establecimiento de tales condiciones, parece necesario recurrir al concepto de *sinonimia* de dudosa claridad[10]. Justamente por este motivo au-

6. Por ejemplo, P. Strawson, 1950, y H. Putnam, 1970.

7. Moretti afirma correctamente que ni siquiera todas las oraciones declarativas tienen condiciones de verdad: «existen usos de oraciones de esa clase, tal vez poco usuales pero legítimos (como en las contraseñas o en las clases de gramática), para los que no tiene sentido hablar de verdad o falsedad de la oración usada» (A. Moretti, 1984, 24).

8. Quizás las oraciones eternas de Quine puedan ayudarnos en este punto: las oraciones de la matemática o aquellas que expresan leyes físicas estarían fuera del alcance de esta objeción.

9. Aunque Moretti argumenta que la existencia de un procedimiento recursivo permite ampliar siempre el conjunto de oraciones-caso efectivas (Moretti, 1984, 28).

10. Raúl Orayen ha argumentado que sin el concepto de *sinonimia* seríamos incapaces de caracterizar algunas de las nociones centrales de la lógica. Véase R. Orayen, 1989, cap. II.

tores como Quine rechazan las proposiciones como legítimos portadores de verdad (Quine, 1970, cap. 1).

Otra razón para rechazarlas tiene que ver con la posibilidad de realizar correctas asignaciones de verdad[11]. Las oraciones tienen estructura. Las proposiciones, al ser objetos extralingüísticos, no parecen tenerla. ¿Cuál es la estructura de una proposición es una cuestión de difícil solución, mientras que decir cuál es la estructura de una oración es una tarea más sencilla. Pero las definiciones de *verdad* y de *verdad lógica* requieren de la estructura de las oraciones. De otra manera, no sería fácil explicar cuál es el valor veritativo de las infinitas fórmulas pertenecientes al lenguaje de la lógica clásica[12].

Ahora bien, tal vez todas estas dificultades con las proposiciones surjan por haber supuesto que tuviéramos que realizar una interpretación objetual de los cuantificadores. Acaso debamos adoptar una lectura sustitucional. Sin embargo, nuevamente, esta opción parece darnos motivos en favor de las oraciones, ya que '(p) (p v ¬p)' se leerá 'todas las instancias de sustitución del esquema '... v ¬ ...'' son verdaderas, donde no parece haber una mejor interpretación que la de que las instancias de sustitución sean entidades lingüísticas: aquellas capaces de ocupar los espacios en blanco del esquema.

Vemos entonces que aunque podamos evitar pensar que la lógica hable sobre proposiciones, no parece posible eludir al menos un compromiso con oraciones tipo. En cualquier caso, no parece que estemos ante la presencia de un tipo de entidad concreta. Y además, conviene recordar que, respecto de nuestros compromisos ontológicos, no sólo tenemos el problema de los portadores de verdad. Un lenguaje para la lógica de predicados contiene constantes individuales (que usualmente son interpretadas como haciendo referencia a una entidad individual u objeto) y constantes predicativas (las que se ven como aquellas expresiones que se aplican o no a ciertos objetos individuales). Pero la pregunta aquí es en virtud de qué se aplican esas expresiones a los objetos individuales correspondientes. ¿A qué nos compromete la representación formal de *Sócrates es mortal*? ¿Por qué se aplica *es mortal* a Sócrates?[13]. Las propiedades y las clases son los candidatos ontológicos a desempeñar este papel. Podemos interpretar que el predicado *es mortal* representa una propiedad. En este caso, lo

11. Quizás convenga recordar en relación con este tema que tanto el propio Tarski como Kripke eligieron la opción oracional en sus trabajos sobre la verdad.

12. Recuérdese la tesis fregeana: el valor veritativo del todo depende del valor veritativo o referencial de las partes.

13. Gonzalo Rodríguez-Pereyra considera que éste justamente es el problema de los universales. Véase Rodríguez-Pereyra, 2000.

que hace verdadera a la oración *Sócrates es mortal* es que Sócrates tenga la propiedad de ser mortal. Pero ¿qué son las propiedades? Muchos son los que piensan que las propiedades son universales. Hay universales inmanentes (aristotélicos)[14] y trascendentes (platónicos). En ambos casos, hay un compromiso con la idea de que las propiedades son «entidades repetibles» (*repeatables*). Los universales trascendentes son existencialmente independientes, mientras que los inmanentes no. Las dos alternativas sostienen que para que Sócrates tenga una propiedad, debe instanciar un universal. Pero, en la primera, el universal es una entidad cuya existencia no necesita de la existencia de ninguna otra entidad. Por supuesto, esto no es lo que pasa con los universales inmanentes. Otro aspecto a explicar es el de qué quiere decir que diferentes particulares tienen una misma propiedad. Como cuando en nuestras explicaciones semánticas decimos que '$p \vee \neg p$' y '$\neg (p \,.\, \neg p)$' comparten alguna propiedad. Desde la perspectiva que estamos considerando, esa frase parece querer decir que ellas instancian un mismo universal. En este enfoque se explica la semejanza entre particulares como una consecuencia de la identidad de los universales que ellos instancian. Dos particulares comparten un universal si se parecen en algún respecto, y si dos particulares se parecen el uno al otro en algún respecto es porque ellos comparten un universal.

Una alternativa distinta es pensar que no hay entidades universales: que para que la oración *Sócrates es mortal* sea verdadera hay que comprometerse con particulares abstractos. Éste es el compromiso de la denominada *teoría de tropos*. Ha sido defendida en diversas variantes por Campbell (1990) y Williams (1953) entre otros. Los defensores de este enfoque concuerdan en que no hay universales: no hay nada que esté completamente localizado en diferentes lugares al mismo tiempo. Pero, más allá de la existencia de entidades concretas, hay particulares abstractos como *la verdad de esta oración,* o *la validez de este argumento,* o *la cardinalidad de este conjunto.*

Para algunos defensores de esta posición, los tropos son el «alfabeto del ser» (*alphabet of being*) (Williams, 1953). Los tropos son las entidades básicas de las que todo lo demás está compuesto. Desde esta perspectiva, los particulares concretos son menos fundamentales

14. Armstrong defiende una teoría de universales inmanentes. Mellor, en cambio, se compromete con universales trascendentes. En contra de estos últimos, se han formulado diversas variantes del argumento aristotélico del tercer hombre. A favor de ellos, se dice que son necesarios para dar cuenta de las posibilidades físicas no realizadas que parecen presuponer las leyes naturales. Véase D. Armstromg, 1978, y D. Mellor, 1995.

que los particulares abstractos. La oración «la nieve es blanca» en tanto particular concreto es menos fundamental que sus tropos: su verdad, su temporalidad. Está constituida de ellos. Los tropos no son repetibles y por ello no son universales. Lo que tienen en común «la nieve es blanca» y «el pasto es verde» no es una tercera entidad la cual es poseída por las dos oraciones. Por el contrario, la verdad de la primera oración es numéricamente diferente de la segunda. Parecen una a la otra no sólo en algún respecto sino exactamente.

Por esta razón, en este enfoque se toma la propiedad de *ser un tropo* y se dice que el que un particular tenga una propiedad depende de que tenga un tropo. Ya que los tropos no pueden ser compartidos por diferentes particulares, para que distintos particulares compartan una propiedad tienen que poseer precisamente tropos semejantes. Por ello «la nieve es blanca» y «el pasto es verde» son ambas oraciones verdaderas, porque tienen tropos que se asemejan el uno al otro. Por esta razón, en lugar de explicar el parecido entre particulares en términos de la identidad de las propiedades que ellos tienen, el enfoque de los tropos la explica en términos de la semejanza de los tropos que ellos tienen (Campbell, 1990, 40). Si dos particulares tienen exactamente tropos semejantes, ellos se asemejan el uno al otro y si ellos se asemejan es porque ellos tienen exactamente algún tropo que se parece. De esta forma, aunque en el enfoque se explique la semejanza entre particulares concretos, trata el parecido entre tropos como un concepto primitivo (Campbell, 1990, 31).

Sin embargo, construir concretos como la suma de tropos está sujeto a la siguiente crítica: si un concreto fuera nada más que una suma de tropos, ¿qué impediría que cualquier suma de tropos, por dispersa e inconexa que sea, determinara un concreto? Hay que poder especificar la relación de *copresencia*. Si esta relación es diádica, será necesario además un pegamento que junte los pares de tropos copresentes entre sí. Además, como sostiene Devitt (1980), no está claro que el defensor de los tropos no adopte la estrategia del avestruz al considerar como primitivo lo que debería ser justificado (la relación de semejanza entre particulares abstractos). No obstante, como señala Rodríguez-Pereyra (2002) el propio Devitt puede ser objeto de la misma acusación (la de adoptar un nominalismo del avestruz). O también, tal como afirma Bordes Solanas (1996, 151), la acusación de Devitt parece circular, ya que para el partidario de los universales sólo está claro que esa relación tiene que ser analizada. En el mismo sentido, el defensor de la teoría de tropos podría objetar al realista de los universales que no ofrece un análisis de la noción de *ejemplificación*. Otra manera de oponerse a los tropos tiene relación con la crítica rus-

selliana a las posiciones nominalistas. Aplicado a los tropos, el argumento se inicia pidiéndonos que consideremos tres tropos: la verdad de la oración *la nieve es blanca*, la de *el pasto es verde* y la de *el cielo es azul*. Cada uno de ellos mantiene con el otro una relación primitiva de semejanza, lo cual justifica que pertenezcan a la misma clase de semejanza. El punto russelliano es que si esta relación no es la misma para todos (no es un universal), entonces entre cada particular y cada otro de los tres ha de haber una relación de semejanza distinta que sea a su vez un tropo. Pero, razonando de manera análoga a la anterior, tales relaciones de semejanza, por ser tropos que relacionan tropos semejantes, serán semejantes a su vez, de modo que, o bien la relación de semejanza entre relaciones de semejanza será lo misma o bien no lo será. Si es la misma, hay universales, si no la es, el argumento puede repetirse hasta el infinito.

Una alternativa al compromiso con universales y los tropos está dada por el nominalismo de semejanza[15]. Según este enfoque las propiedades de los particulares se explican en términos de las clases de particulares que se asemejan entre sí. La semejanza entre particulares es algo básico que no puede ser reducido a nada más elemental. En particular, lo que se niega es que la semejanza entre particulares pueda ser entendida en términos de una característica común a ellos. Eso es lo que justamente haría un realista acerca de los universales. De esta manera, para que la oración *la nieve es blanca* sea verdadera debe parecerse a todos los otros particulares que son verdaderos. El parecido entre particulares concretos no es explicado, pero se usa para explicar las propiedades de estos particulares. Es el parecido entre particulares concretos lo básico y fundamental, y no, como en la teoría de tropos, el parecido entre particulares abstractos. En suma, los nominalistas aceptan la existencia de propiedades, pero no la de entidades universales. Las propiedades son clases (maximales) de particulares que se asemejan entre sí.

El nominalismo de semejanza tampoco carece de problemas. Uno de ellos es que esa posición implica adoptar una teoría relacional de las propiedades, ya que la semejanza es una relación, y para que un particular tenga cualquier propiedad tiene que parecerse a otros particulares. Por supuesto, a diferencia de otras teorías relacionales, los *relata* que sirven para dar cuenta de las propiedades son todos parti-

15. Armstrong identifica distintas versiones nominalistas: entre las principales encontramos el llamado *nominalismo del avestruz* (*Ostrich Nominalism*) de Quine, Devitt, Aune y van Cleve, el nominalismo predicativo o conceptual, el nominalismo mereológico, el nominalismo de clases de Quinton y D. Lewis, y el nominalismo de semejanza. Véase D. Armstrong, 1978, 12-17.

culares concretos. Pero, van Cleve (1994, 580) sostiene que ninguna teoría relacional funciona. La dificultad de la coextensión formulada contra el nominalismo de semejanza (Armstrong, 1989, 50-51, y 1997, 162), surge de tomar en cuenta casos de propiedades coextensivas: propiedades que tengan exactamente los mismos particulares. Si el nominalista de semejanzas identifica las propiedades con clases, la dificultad que surge es evidente: la corrección del nominalismo de semejanza implica que no puede haber propiedades coextensivas.

El nominalismo de semejanzas no sólo admite particulares concretos, admite clases. Las clases, como Armstrong reconoce (1997, 188), son también particulares, ya que no son repetibles ni instanciables. Por supuesto, se diferencian de los particulares concretos en ser *abstractas*: ellas son particulares abstractos. Por esta razón, el nominalismo de semejanza tiene una ontología monista que se compromete únicamente con particulares (concretos y abstractos). El realismo admite un tipo extra de entidad (los universales abstractos). Sin embargo, lo que el nominalismo de semejanza gana eliminando de su ontología a los universales, el realismo lo obtiene reduciendo las clases a particulares y universales.

De todas formas, para que la oración «Sócrates es mortal» sea verdadera podemos no tomar en cuenta las propiedades mismas, sino las clases de los objetos que las tienen[16]. Esta opción tiene la ventaja de enfatizar que, en el lenguaje de la lógica de predicados, para determinar el valor de verdad de una fórmula lo único que necesitamos saber es cuáles de las cosas del dominio tienen esa propiedad. Como sostiene Gamut, «(n)o importa, por ejemplo, cómo lo sabemos ni si las cosas podrían ser diferentes» (1991, 90). Sin embargo, aunque las clases tengan, a diferencia de las propiedades, condiciones de identidad claras[17], la idea intuitiva de hablar de colecciones de entidades de cualquier tipo puede traernos sorpresas. Son conocidos los problemas relacionados con las colecciones infinitas «demasiado grandes» y las paradojas que parecen involucrar (véase Sartorio, 2000).

Aunque dejemos a un lado el problema de los portadores de verdad y sus componentes estructurales y abordemos directamente los argumentos, desde un punto de vista puramente sintáctico, si lo que nos importa es ofrecer un criterio de corrección, nos vemos obligados a recurrir a los conceptos de *regla de inferencia* y de *prueba*. En efec-

16. Por supuesto, en el caso de que estemos tratando de interpretar un predicado monádico.

17. Si dos clases tienen exactamente los mismos elementos, entonces son la misma clase.

to, de lo que se trata cuando nos preguntamos por la corrección de los argumentos particulares es de dar una lista pequeña y siempre aparentemente trivial de pasos que se justifican en la aplicación de reglas inferenciales. Es usual llamar a esas cadenas de fórmulas *derivaciones formales*. A la última línea de la secuencia se la llama *conclusión*. Y a aquellas líneas que no requieren justificación, *premisas*. Por supuesto, es posible incluir en la derivación alguna línea sin justificación. A esa línea se la llama *supuesto*. Pero, esa inclusión está formalmente permitida sólo si *cancelamos* ese supuesto mediante la utilización de una regla. Teniendo en cuenta lo anterior, se dirá que una inferencia de la forma ∅1, ..., ∅n / ψ es formalmente correcta si y sólo si hay una derivación de ψ desde ∅1, ..., ∅n. Un sistema que adopte este enfoque recibe el nombre de *deducción natural*. Como hemos dicho, es central en este enfoque el recurso a la noción de *regla de inferencia*, ya que de otra manera no se entiende qué es una prueba. Por ello los sistemas de deducción natural se presentan como sistemas de reglas de inferencia. El hecho de que ψ se siga de $\emptyset_1$, ..., $\emptyset_n$ o que $\emptyset_1$, ..., $\emptyset_n$ / ψ sea formalmente correcto indica que hay una o varias reglas de inferencia que autorizan el paso de $\emptyset_1$, ..., $\emptyset_n$ a ψ. Pero son muchos los que dudan acerca de que haya elementos objetivos que garanticen la corrección de una secuencia[18]. Si las reglas fueran entidades lingüísticas, si ellas fueran solamente las oraciones que las expresan, estarían sujetas a diversas interpretaciones, algunas de las cuales serían completamente inaceptables. Pero si ellas no son entidades lingüísticas, ¿qué tipo de entidades son?

Podría pensarse que el recurso a entidades no concretas puede ser evitado si se adopta un enfoque semántico[19]. Este pensamiento también podría justificarse en que por lo general los métodos semánticos tienden a dar una mejor comprensión que los métodos puramente sintácticos no sólo por su alto poder heurístico sino por sus frecuentes aplicaciones al estudio del lenguaje natural. Sin embargo, es fácil advertir que muy pronto volvemos a encontrar entidades aparentemente oscuras. En efecto, si lo que nos importa es la validez de una inferencia del tipo $\emptyset_1$, ..., $\emptyset_n$ / ψ, será válida si y sólo si no hay un modelo M en el cual al interpretar todas las letras de predicados y constantes lógicas que aparezcan en $\emptyset_1$, ... , $\emptyset_n$, / ψ, la valuación del conjunto $\{\emptyset_1, ..., \emptyset_n\}$ resulte verdadera y resulte falsa ψ. Es obvio que la definición semántica de *validez* está basada sobre una miste-

18. Wittgenstein, 1958, y Kripke, 1982.
19. Para una comparación entre el enfoque sintáctico y semántico de la noción de consecuencia lógica, véase Alchourrón, 1995.

riosa totalidad: la clase de todos los modelos. Hay infinitos modelos y cada uno puede ser infinitamente grande. Además, por supuesto, los modelos son entidades conjuntistas. Y como hemos afirmado anteriormente, las entidades de este tipo no son entidades concretas.

En el ámbito de las extensiones de la lógica clásica los compromisos con oscuras entidades también abundan. Por citar un caso, reflexionemos sobre las lógicas modales. En las semánticas de este tipo de lógicas se relativiza la noción de *verdad* a entidades llamadas *mundos posibles*. El procedimiento es simple: se incorpora, al menos, otro elemento (otro conjunto) en la estructura del modelo (Orayen, 1995). Está claro que la semántica de las modalidades ha mostrado que el concepto de *mundo posible* tiene un considerable poder heurístico. Al proponer que la necesidad (o la posibilidad) sea interpretada en términos de cuantificadores que recorren mundos, se obtiene una interpretación extensional de las mencionadas expresiones intensionales. Estos desarrollos han aportado mucho, sin duda, al establecimiento de las condiciones veritativas de estas expresiones y por ende a la comprensión de las inferencias en que están involucradas. Sin embargo, no es sencillo precisar qué son los mundos posibles. Al respecto, dice Lewis:

> Los mundos son como planetas remotos; excepto en que la mayor parte de ellos son mucho más grandes que los meros planetas y en que ellos no son remotos. [...] Ellos no están a ninguna distancia de aquí. Ellos no están tampoco ni en el pasado ni en el futuro [...] (Lewis, 1986, 2).

No tenemos ni relaciones causales ni espacio-temporales con tales cosas. Por ello, los mundos deben ser tomados como entidades aisladas (*isolated*). Cada uno de los mundos está causalmente clausurado y a ninguna distancia espacio-temporal del otro. Ellos *no están ni en el pasado ni en el futuro, ni a ninguna distancia lejana de aquí*[20].

La posición de Lewis no es la única alternativa. Stalnaker (1976), por ejemplo, sostiene que las modalidades hacen referencia sólo a cosas pertenecientes a nuestro mundo. Las posibilidades son modos no instanciados: alternativas que nuestro mundo podría llegar a ser. La

20. A pesar de lo anterior, Lewis argumenta que los mundos son particulares concretos. Sostiene que del mismo modo en que no creemos que nuestro mundo sea una entidad lingüística o conceptual, tampoco parece razonable que los restantes mundos a los que hacemos referencia con nuestro discurso modal, sean entidades de este tipo. Estrictamente hablando, los mundos posibles y sus partes no son, cosas que podrían llegar a ser, ni entidades abstractas.

idea de un *modo no instanciado* remite a la de *propiedad* y, por supuesto, a los problemas involucrados con ella. Por supuesto, también hay interpretaciones lingüísticas de las modalidades[21]. Pero no parecen totalmente convincentes (véase Lewis, 1986, cap. 3). Otra cuestión surge al elaborar una semántica para nuestras oraciones modales, la de la naturaleza de la relación entre los mundos del modelo. Decimos que desde nuestro mundo podemos acceder a aquel mundo posible y que desde aquél podemos acceder a otro mundo alternativo. Eso que parecen compartir esas dos afirmaciones puede llevarnos a pensar que hay una entidad que ambas comparten. ¿Habrá una relación de accesibilidad? Reaparecen entonces, aunque tratemos de evitarlas, cuestiones relacionadas con el problema de la identidad genérica.

En suma, la comprensión literal de las oraciones de la lógica parece involucrar un compromiso con entidades cuya naturaleza es algo oscura. Quizás podamos evitar comprometernos con proposiciones, quizás podamos evitar las propiedades a través de las clases. Pero aun así persistirán cuestiones emparentadas con entidades con las que no compartimos el espacio tiempo, con las que no estamos en contacto causal, ni forman parte de los objetos que fácilmente podemos estar dispuestos a aceptar.

2. ¿QUÉ SON LAS ENTIDADES ABSTRACTAS?

El deseo de encontrar criterios que nos permitan identificar entidades abstractas no puede fácilmente ser satisfecho. Uno de los que se han ofrecido se relaciona con la posibilidad de advertir empíricamente su existencia. De acuerdo con este criterio, las *entidades concretas* son aquellas cuya existencia podemos experimentar. Las *entidades abstractas,* por el contrario, carecen de esta característica. Ellas, si las hay, no pueden ser vistas ni tocadas. Podemos ver esta fórmula o la verdad de esta oración, pero no las vemos como entidades abstractas. Este primer criterio, sin embargo, parece fallar en tanto intento para establecer una supuesta distinción ontológica entre entidades. Es claro que no parece aceptable recurrir a la inaccesibilidad experimental porque ese recurso contiene una indeseable relatividad a las facultades humanas. Estamos tratando de establecer una distinción ontológica y este criterio no parece aceptable para tal establecimiento. Además, no es muy claro cuál es el alcance de nuestras facultades perceptuales. No es sencillo decir si nos estamos limitando a lo que podemos ver, sentir

21. Como por ejemplo la de R. Carnal, 1947.

y tocar o también debemos incluir lo que podemos percibir a través de instrumentos. Parece natural incluir estos últimos dentro de los concretos. Pero, hasta tanto no exista tal instrumento, ¿estaríamos dispuestos a sostener que tales entidades son abstractas?

Una alternativa al anterior criterio es sostener que las *entidades abstractas,* a diferencia de las *concretas,* carecen de una locación espacio-temporal. De acuerdo con este criterio, *la propiedad de ser una verdad lógica* no parece ser una característica que una fórmula tenga sólo en este momento. Por supuesto, en este momento la tiene. Pero será la misma característica que poseerá más tarde y que poseyó anteriormente. El punto es que no parece tener sentido preguntarse si la misma característica la tuvo ayer y la tiene hoy. En la misma línea, tampoco parece tener completo sentido preguntarse por el lugar que ocupa en el espacio una entidad abstracta. Una característica abstracta puede estar totalmente presente en más de una entidad concreta al mismo tiempo y en distinto lugar. El problema con esta posición es, no obstante, que hay objetos concretos que no están localizados ni espacial ni temporalmente. Un fantasma, o los personajes de una novela parecen ser entidades concretas sin localización real. Además, puede pensarse que, aunque abstractas, algunas clases tienen localización[22]. Y hay características abstractas que se pueden adquirir en un momento determinado[23].

Otra alternativa es sostener que los *objetos concretos*, a diferencia de los *abstractos*, son capaces de moverse o de desarrollar cambios intrínsecos. Sin embargo, nuevamente es posible encontrar objetos concretos que no cumplen esta nota. Como argumenta Rosenkrantz (1995), los instantes temporales son concretos[24], pero son incapaces de moverse.

También podemos buscar un criterio de distinción entre lo concreto y lo abstracto en el tipo de existencia: mientras que los *concretos* tendrían una existencia contingente, los *abstractos* tendrían una existencia necesaria. El problema nuevamente es que intuitivamente tratamos a algunos objetos concretos como si tuvieran una existencia

22. Contra este punto de vista argumenta Joan Pagès (2000).

23. Además, podría pensarse con Armstrong que hay particulares abstractos y que son metafísicamente dependientes del espacio-tiempo que ocupan. Ellos no conservarían su identidad si no se hubieran dado en el lugar y el tiempo en que se dieron. En este sentido, no hay más particulares concretos que las regiones espacio-temporales y los particulares abstractos son algo así como propiedades particulares suyas (Armstrong, 1989).

24. Para una discusión sobre la ontología de los momentos e intervalos temporales en lógica temporal, véase R. Morado, 1998.

necesaria. No parece razonable decir que Dios, si es que existe, es un objeto abstracto.

Una alternativa a todo lo anterior es decir que las entidades abstractas son ejemplificables mientras que las concretas no. En la misma línea, se podría afirmar que las *entidades abstractas* tienen que ser literalmente idénticas en cada una de sus instancias mientras que las *concretas* sólo son idénticas a sí mismas. La objeción aquí es que podríamos pensar que hay ciertos conjuntos, proposiciones o propiedades que no podrían estar ejemplificados (tales como la propiedad de *ser un cuadrado redondo*). En este caso, estaríamos en presencia de entidades abstractas que no pueden estar ejemplificadas.

Otra alternativa es pensar que los *objetos concretos* son aquellos a los que podemos señalar ostensivamente, mientras que los *abstractos* no. Pero, nuevamente, este criterio parece caer en una capacidad humana que no permitiría trazar una distinción metafísica. Recuérdense, además, al respecto los problemas señalados por Wittgenstein a la ostensión (1958, 27 ss.).

Finalmente, podríamos inclinarnos a pensar que las *entidades concretas* son independientes y completas, mientras que las *entidades abstractas* son dependientes e incompletas (Campbell, 1990, 2-3). Las cosas que ni se dicen ni están en un sujeto (como este lápiz) son concretas, las que se dicen de un sujeto (como la negrura) son abstractas. El punto para sostener este criterio como adecuado es entender qué quiere decir *metafísicamente dependiente*. Como sostiene Bordes Solanas, «la relación de dependencia metafísica es complicada de entender» (1996, 140). Esta dependencia parece ser sin duda necesaria y asimétrica y darse en un entramado *a priori* (entre cosas que no suelen estar unidas en la experiencia). Pero lo primero que hay que advertir es que esta relación no parece ser una relación mereológica. Sólo los concretos son partes de concretos. Esta relación, por ser necesaria, tampoco es una relación de dependencia contingente. No parece ser una cuestión empírica sino conceptual. Obviamente, parece natural sostener que la verdad de la oración que estoy escribiendo depende metafísicamente de esta oración, porque sin ella no existiría. Pero ¿depende la identidad de los abstractos de la de los concretos? ¿Y en qué sentido si lo hace? Otro modo de entender lo subordinado o secundario lo es por su dependencia epistemológica. Pero esta alternativa estaría sujeta a la misma objeción que la que hicimos cuando tratamos la idea de estar más allá de nuestras capacidades experimentales. Se dirá que los objetos primarios de conocimiento parecen ser los concretos y a lo abstracto parece llegarse por un esfuerzo cognitivo de abstracción que consiste en restar detalles a lo concreto. El

conocimiento de las entidades abstractas depende del conocimiento previo de otra cosa (concreta) y es incompleto. Pero como ya hemos visto, este sentido de dependencia, sin embargo, no fija de modo absoluto la subordinación de unas cosas a otras, sino sólo de modo relativo a los seres humanos. Es posible que las cosas abstractas sean secundarias epistemológicamente, pero constituyen los cimentos últimos de la realidad.

En síntesis, más allá de que tengamos la distinción intuitiva entre entidades concretas y abstractas, no resulta sencillo encontrar un criterio que la establezca. Quizás debamos abandonar tal pretensión. Pero hagamos lo que hagamos, ni las oraciones tipo, ni las clases, ni las propiedades, ni esas clases especiales que llamamos modelos, ni las reglas, ni las relaciones de accesibilidad entre mundos parecen ser entidades concretas. ¿Qué significa comprometernos ontológicamente con ellas? ¿Qué razones tendríamos para realizar ese compromiso?

3. LA DISPUTA ONTOLÓGICA ACERCA DE LAS ENTIDADES ABSTRACTAS

La lógica parece hablar, en parte, de entidades que están más allá de lo concreto. El entenderla literalmente, el no tratarla como un discurso de ficción, nos compromete con entidades cuya naturaleza parece ser distinta a la de las entidades concretas. Adoptar el realismo frente a esas entidades es tomar una actitud ontológica. De acuerdo con esa actitud, las clases, propiedades, reglas y funciones tienen existencia real, una existencia que es completamente independiente de nosotros, de nuestras formas de hablar, actuar y pensar. Ellas no tienen una naturaleza mental o lingüística.

Por supuesto, hay diversos matices dentro del realismo. El extremo o platónico sostiene que las entidades abstractas trascienden lo particular. Ellas sólo pueden ser conocidas por el intelecto, y no por los sentidos. Generalmente, en esta concepción, para tener conocimiento de lo que es concreto necesitamos acceder al universal inalterable. Los concretos son sólo manifestaciones de las formas abstractas. La actitud más moderada asociada al aristotelismo considera que las entidades abstractas no existen independientemente de las concretas. Lo abstracto sólo está presente en las entidades concretas que encontramos en nuestra experiencia. Ellas existen independientemente de los pensamientos humanos, pero no en el mismo sentido que las entidades concretas como los objetos físicos. Se encuentran en los atributos compartidos específicos de los objetos particulares. Por ejem-

plo, el objeto abstracto *blancura* se encuentra en la clase de todos los objetos blancos individuales, tales como las nubes y el algodón.

También podemos adoptar una actitud ontológica antirrealista. En esta opción, en cambio, se rechaza la existencia de las entidades abstractas. Es usual que los antirrealistas conecten su actitud con planteos de índole epistemológica o semántica. En la primera línea, objetan la postulación de objetos abstractos por tratarse de objetos que parecen estar más allá de nuestras capacidades cognitivas. No hay contacto causal con los objetos abstractos y ello hace que sea difícil explicar cómo los conocemos (Benacerraf, 1973). En la otra línea, se cuestiona al realismo en el plano semántico: la explicación que éste da de la comprensión de las oraciones que parecen involucrar entidades abstractas se sustenta en la atribución a los usuarios del lenguaje de una capacidad que trasciende sus capacidades efectivas (Dummett, 1978). Los dos argumentos se unen: ambos rechazan la admisibilidad de la tesis realista según la cual hay entidades abstractas acerca de las que hablan las oraciones de la lógica que están más allá de nuestras capacidades de reconocimiento o mecanismos epistémicos de captación. ¿Por qué son válidas las leyes de la lógica clásica? La actitud antirrealista liga la respuesta a esta pregunta con nuestra captación del significado de las constantes lógicas. Los que simpatizan con esta actitud sostienen que, en el plano semántico, nuestra captación de ese significado debe ser plenamente manifestable en nuestra conducta inferencial. En el epistémico, afirman que nuestra captación de las entidades presuntamente involucradas no debe incluir la posesión de capacidades que nos trasciendan. Ambas ideas objetan la plausibilidad de que la corrección de las oraciones de la lógica consista en que posean condiciones veritativas que nos trasciendan (tal como, según ellos creen, dice todo realista).

Más allá de la evaluación que hagamos de esta polémica entre realistas y antirrealistas, existe otra línea de argumentación que relaciona la aceptación de las entidades abstractas con el éxito de las teorías en las cuales resulta inevitable su presencia. Según esta línea de argumentación, las entidades abstractas son indispensables para hacer lógica, metalógica y hasta ciencia empírica[25]. La estrategia de la indispensabilidad no argumenta directamente a favor de la existencia de entidades abstractas. Se concentra, en cambio, en el papel que cum-

25. El argumento de la indispensabildad de las entidades abstractas aparece en varios escritos de Quine y Putnam. Marco Ruffino me llamó la atención sobre algunas tempranas versiones del mismo en los trabajos de Frege, Zermelo y Gödel. Quine, 1948 y 1954, en H. Putnam, 1967, 1970 y 1975.

plen dichas entidades en el resto de la ciencia como un todo. Sostiene Quine (1948):

> Nuestra aceptación de una ontología es, pienso, similar en principio a nuestra aceptación de una teoría científica [...]: adoptamos, por lo menos en la medida en que seamos razonables, el esquema conceptual más simple en el cual los fragmentos desordenados de experiencia bruta puedan ser ajustadas y organizadas [...] Una ontología platonista es, desde el punto de vista de un esquema conceptual fisicalista, tan mítica como lo es un esquema conceptual fisicalista desde la perspectiva fenomenalista. Este mito superior es bueno y útil, en la medida en que simplifique nuestra explicación de lo físico.

En esta estrategia la aceptación de objetos abstractos es similar a la aceptación de otros tipos de objetos postulados por las ciencias empíricas. Del mismo modo que aceptamos una ontología fisicalista en lugar de una fenomenalista y justificamos esta aceptación sobre la base de que con ella logramos una mayor simplicidad en nuestra explicación global del mundo observable, la aceptación de ciertas entidades abstractas está también justificada por la mayor simplicidad que con su aceptación se logra tomando la ciencia como un todo.

La suposición más controvertible del argumento es aquella según la cual la cuestión acerca de lo que hay debe ser respondida observando nuestras mejores teorías científicas. Sólo reconoceremos como existentes aquellas cosas que son posibles valores de las variables de nuestras mejores teorías científicas. Ya que ciertas entidades abstractas forman parte del rango de las variables usadas en lógica, estas entidades abstractas existen. Por supuesto, como el propio Quine hace notar, esta conclusión no nos limita la posibilidad de cambiar nuestros pensamientos al respecto en el futuro. Si encontramos una teoría mejor que sea radicalmente diferente a la nuestra, no sólo en su contenido específico sino en la lógica subyacente que se usa en su estructuración, y que cuantifica sobre entidades completamente distintas, nuestros compromisos deberán cambiar. En suma, la única teoría lógica que resulta aceptable y que por ello hay que adoptar todos sus compromisos ontológicos es aquella que en yuxtaposición con una teoría científica posee alguna aplicación empírica.

Uno de los problemas con el argumento de la indispensabilidad es que nuevamente supone una subordinación de la ontología a la epistemología. Claro que, en este caso, no estamos tratando de establecer un criterio sino más bien encontrar una buena razón para dicha aceptación. Y quizás podamos confiar en nuestro éxito empírico. Claro que para que tenga sentido esta línea argumentativa en la ló-

gica debemos también confiar en el holismo confirmacional y en el rechazo de la distinción analítico-sintético (tal como lo hace el propio Quine).

Otro problema es que sólo una parte de la lógica y de la matemática encuentra alguna aplicación en la ciencia empírica. En esta línea, Ruffino nos pide que prestemos atención a lo que pasa con la teoría de conjuntos (2001, 257). Un argumento similar puede ser encontrado en Parsons (1986, 382). Usualmente se supone la existencia de conjuntos de una cardinalidad mucho más grande que aleph-0 o aleph-1 (las cuales son las cardinalidades del conjunto de los números naturales y de los números reales respectivamente). Ninguna aplicación de estos conjuntos puede ser encontrada en nuestro espacio-tiempo. Parece seguirse del argumento de la indispensabilidad que no hay justificación para la existencia de conjuntos con estas cardinalidades, lo cual parece implicar en forma derivada que sólo los números naturales y racionales son indispensables para la ciencia, ya que por la prueba de Cantor ellos tienen una cardinalidad que es compatible con las aplicaciones de nuestras teorías empíricas.

Por supuesto las personas que trabajan en teoría de conjuntos no proceden de este modo. Como el propio Parsons señala, ni el propio Quine, cuando trabaja en teoría de conjuntos, procede de esta manera. El punto no parece preocupar a Quine en absoluto. En su réplica a Parsons dice lo siguiente:

> La matemática pura, en mi punto de vista, está firmemente insertada como una parte integral de nuestro sistema del mundo. Por eso, mi enfoque de la matemática pura está orientado estrictamente a la aplicación en ciencia empírica [...] Tenemos un modesto ejemplo en el proceso con los números irracionales: ninguna medida podía ser adecuada para estar acomodada a los números racionales, si no admitimos simplificar nuestros cálculos y generalizaciones. Con la teoría de conjuntos pasa lo mismo. Reconozco infinitos no numerables sólo porque estoy forzado por las sistematizaciones cognitivas más simples de asuntos bienvenidos. Las magnitudes en exceso de tales demandas o los números inaccesibles, los veo sólo como un pasatiempo matemático sin derechos ontológicos (1986, 382).

Esto es decir que únicamente estamos autorizados a afirmar la existencia de las entidades postuladas por la teoría de conjuntos cuando ellas cumplen un papel para organizar otros tipos de entidades menos abstractas que encuentran una directa aplicación en las ciencias empíricas. Todas aquellas entidades que no cumplan con este propósito (por ejemplo, los cardinales inaccesibles) no tienen que ser reco-

nocidas como existentes. Por supuesto, Quine no afirma explícitamente que la existencia de estas entidades deba ser rechazada. Pero si se tiene en cuenta su doctrina según la cual las entidades existentes son sólo aquellas cuya existencia es requerida por teorías exitosas, estamos muy cerca de este rechazo.

Otra famosa crítica al argumento de la indispensabilidad es la formulada por Elliot Sober (1993). Sober encuentra una diferencia fundamental entre apelar a la indispensabilidad de la existencia de las entidades teóricas de la ciencia empírica y apelar a la indispensabilidad de los objetos abstractos supuestos por la lógica. En el primer caso, esas entidades tienen ventajas explicativo-predictivas con respecto a sus teorías rivales. Pero, de acuerdo con Sober, en el caso de las entidades abstractas no tiene sentido hablar de teorías rivales, ya que toda teoría empírica presupone la matemática y la lógica. La idea misma de *teoría rival* presupone que respecto del mismo fenómeno, cuando una es verdadera, la otra será falsa.

4. CONSIDERACIONES FINALES

Hemos visto que la lógica parece hablar acerca de entidades cuya naturaleza está más allá de lo concreto. Parece difícil permanecer en este plano, cuando necesitamos fórmulas capaces de expresar infinitas cosas, predicados que nos comprometan con atributos, clases o tropos, reglas de inferencias y secuencias de fórmulas, funciones y modelos, mundos y relaciones de accesibilidad. Sin embargo, he argumentado que a pesar de esta motivación, no resulta fácil ofrecer un criterio que permita determinar las condiciones que debe cumplir una entidad para ser considerada como concreta o abstracta. Tenemos la distinción intuitiva, pero parece complicado abandonar este plano. Finalmente, he ofrecido uno de los principales argumentos a favor del compromiso con entidades abstractas. Éste es conocido como el *argumento de la indispensabilidad.* El argumento supone que todo abandono del plano de lo concreto es sospechoso. A partir de lo cual se afirma que el paso de un plano a otro está justificado cuando es indispensable postular algo abstracto como parte de una teoría científica exitosa. El rechazo de este argumento, sea por las razones ofrecidas por Ruffino, Parsons o Sober, o por otras, nos deja con la necesidad de encontrar límites a la mencionada postulación, si nuestros gustos ontológicos se inclinan por la sospecha frente a lo abstracto.

BIBLIOGRAFÍA

Nota: los libros y artículos incluidos en esta bibliografía son, exclusivamente, los citados a lo largo del texto.

Alchourrón, C. (1995), «Concepciones de la lógica», en C. Alchourrón, J. Méndez y R. Orayen, *Enciclopedia Iberoamericana de filosofía*, vol. 7: *Lógica*, Trotta, Madrid.

Armstromg, D. (1978), *Universals and Scientifics Realism*, Cambridge, Cambridge University Press.

Armstromg, D. (1989), *A combinatorial Theory of Possibility*, Cambridge, Cambridge University Press.

Armstrong, D. (1997), *A World of States of Affairs*, Cambridge, Cambridge University Press.

Benacerraf, P. (1973), «Mathematical Truth»: *J. of Phil.* LXX.

Bordes Solanas, M. (1996), «Primacía y dependencia metafísica: consideraciones acerca de las ontologías de tropos»: *Análisis Filosófico* XVI, 2.

Campbell, K. (1990), *Abstract Particulars*, Blackwell, Oxford.

Carnal, R. (1947), *Meaning and Necessity*, The University of Chicago Press, Chicago.

Cleve, J. van (1994), «Predication without Universals? A Fling with Ostrich Nominalism»: *Philosophy and Phen. Research* LIV, 577-90, esp. 580.

Devitt, M. (1980), «"Ostrish Nominalism" or "Mirage Realism"?»: *Pacific Phil. Quarterly* 61.

Dummett, M. (1978), *Truth and others enigmas*, Harvard University Press, Massachusetts.

Gamut, L. T. F. (1991), *Introduction to Logic*, The University of Chicago Press, Chicago.

Kripke, S. (1982), *Wittgenstein On Rules and Private Language*, Harvard University Press, Cambridge.

Lewis, D. (1986), *On the Plurality of World*, Basil Blackwell, Oxford.

Mellor, D. (1995), *The Facts of Causation*, Routledge, London.

Morado, R. (1998), «Momentos e intervalos: problemas filosóficos en lógica temporal»: *Análisis Filosófico* XVIII, 1.

Moretti, A. (1984), «Los objetos de predicación de verdad»: *Análisis Filosófico* IV, n.º 2.

Orayen, R. (1989), *Lógica, significado y Ontología*, UNAM, México.

Orayen, R. (1995), «Lógica Modal», en C. Alchourrón, J. Méndez y R. Orayen, *Enciclopedia Iberoamericana de filosofía*, vol. 7: *Lógica*, Trotta, Madrid.

Pagès, J. (2000), «Universales y sustratos»: *Análisis Filosófico* XX, 1 y 2.

Parsons, Ch. (1986), «Quine on the Philosophy of Mathematics», en Hahn y Schilpp (eds.), *The Philosophy of W. V. Quine*, Open Court, La Salle.

Putnam, H. (1967), «Mathematics without Foundations»: *J. Of Phil.* LXIV.

Putnam, H. (1970), *Philosophy of Logic*, Harper Torchbooks, New York.

Putnam, H. (1975), «What is Mathematical Truth?», en *Mathematics, Matter and Method*, Cambridge University Press, Cambridge.

Quine, W. (1948), «On What There Is», reimpreso en W. Quine, *From a Logical Point of View*, Harper Torchbooks, New York, 1953).
Quine, W. (1953), *From a Logical Point of View*, Harper Torchbooks, New York.
Quine, W. (1966), «Carnap on Logical Truth» (1954), en W. Quine, *The Ways of Paradox and other Essays*, Random House, New York.
Quine, W., *Philosophy of Logia*, cap. 1, Harvard University Press, Cambridge.
Rodríguez-Pereyra, G. (2000), «What is the Problem of Universals?»: *Mind*, 109.
Rodríguez-Pereyra, G. (2002), *Resemblance Nominalism. A Solution to the Problem of Universals*, Oxford University Press, Oxford.
Rosenkrantz, G. (1995), «Concrete/abstract», en J. Kim y E. Sosa, *A Companion to Metaphysics*, Blackwell, Massachusetts.
Rufino, M. (2001), «Some Remarks about Indespensability Arguments», en U. Pinheiro, M. Ruffino y P. Junqueira Smith, *Ontologia, Conhecimento e Linguagem*, Mauad, Río de Janeiro.
Sartorio, A. (2000), *Conjuntos e Infinitos*, EUDEBA, Buenos Aires.
Sober, E. (1993), «Mathematics and Indispensability»: *Phil. Review*, 102.
Stalnaker, R. (1979), «Actualism and possible world», en M. Loux, *The Possible and the Actual*, Cornell University Press.
Strawson, P. (1950), «On referring»: *Mind*, LIX.
Williams, D. (1953), «On the elements of being I»: *Review of Metaphysics*, 7.
Williams, D. (1953), «On the elements of being II»: *Review of Metaphysics*, 7.
Wittgenstein, L. (1958), *Philosophical Investigation*, Basil Blackwell, Oxford.

TIPOS LÓGICOS, LENGUAJE Y FILOSOFÍA

Francisco Rodríguez Consuegra

Russell llegó a su célebre paradoja de las clases reflexionando sobre la tesis de Cantor de que no existe el número cardinal máximo. Esto puede probarse partiendo de que, si una clase tiene n miembros, entonces hay 2^n formas de combinar esos miembros (como se ve fácilmente ya para los números finitos), y de que el número de subclases de una clase dada es siempre mayor que el número de sus miembros, lo que fue explícitamente probado por Cantor en 1892.

Sin embargo, puede llegarse a una contradicción con sólo conceder que las subclases de una clase dada puedan también ser sus miembros. Así, la clase de todos los objetos debe tener el número cardinal máximo, pues por muy alto que sea el número de las subclases contenidas en ella, ese número no puede superar al de sus miembros, por lo ya concedido. Por tanto, parece que el número cardinal máximo exista y no exista al mismo tiempo. Alternativamente, podría también argumentarse que algo falla en la naturaleza de las clases mismas: podría incluso pensarse que la clase de todos los objetos no pueda tener a las subclases contenidas en ella como miembros, con sólo negar a las clases el estatuto ontológico de objetos. Un tal recurso conduciría de hecho a una teoría «sin clases», resultado que formó parte esencial de la lógica russelliana en su forma más elaborada, pero lo que importa ahora es ver cómo Russell se vio conducido a su famosa paradoja partiendo de la reflexión anterior.

La clase de *todos* los objetos, al ser absolutamente omniabarcante, debe tenerse a sí misma como un miembro más, con lo que al menos esa clase debe ser miembro de sí misma. Otras clases podrían también ser miembros de sí mismas, como sucede con la clase de todas las clases con más de tres miembros: puesto que esa clase tendría, a su

vez, más de tres miembros (hay más de tres clases con más de tres miembros), debería estar contenida en ella misma a título de un miembro más. Se ve, sin embargo, que para que una clase sea miembro de sí misma, debe ser una clase de clases de objetos, y no simplemente una clase de objetos que no sean a su vez clases. Así, una clase de objetos cualesquiera, por ejemplo, de ladrillos, no puede ser miembro de sí misma, pues no es un ladrillo.

Si hay clases que son miembros de sí mismas y clases que no lo son, podremos sin duda agruparlas en dos clases: la de todas las clases que son miembros de sí mismas y la de todas las clases que no son miembros de sí mismas. Consideremos, por mor del argumento original de Russell, la clase de todas las clases que no son miembros de sí mismas, y preguntémonos: ¿es esa clase miembro de sí misma? Es decir, ¿es la clase de todas las clases que no son miembros de sí mismas miembro de sí misma o no? Parece que debe haber una respuesta correcta, sea afirmativa o negativa.

Pero veamos: si respondemos que sí, entonces es que la clase de todas las clases que no son miembros de sí mismas es miembro de sí misma, con lo que llegamos a que esa clase no es miembro de sí misma. Efectivamente, si es miembro de sí misma, debe cumplir la propiedad esencial de todos sus miembros: la de ser una clase que no sea miembro de sí misma. Así, si es un miembro de sí misma, entonces es que no lo es.

Respondamos, pues, que no, con lo que la clase de todas las clases que no son miembros de sí mismas no será un miembro de sí misma. Pero hacerlo así conducirá a que esa clase sí fuera un miembro de sí misma. En efecto, si no es un miembro de sí misma, debe inclumplir la propiedad esencial de todos sus miembros: la de ser una clase que no sea miembro de sí misma. Luego al responder que no es un miembro de sí misma lo que hacemos es reconocer que cumple esa propiedad esencial. Así, si no es un miembro de sí misma, entonces es que sí lo es. Ambas respuestas conducen, pues, a su contraria: estamos ante la paradoja de Russell.

El núcleo de la teoría de los tipos lógicos, como solución de esta paradoja, consiste en denunciar cualquier expresión en la que se diga que una clase es, o pueda ser, miembro de sí misma, como expresión carente de sentido, sobre la base de que una clase y sus miembros pertenecen a tipos distintos. Así, el símbolo de pertenencia reclama, a su derecha y a su izquierda, una diferencia en la jerarquía de tipos, de modo que lo situado a la izquierda sea de un tipo «inferior» a lo situado a la derecha, pero nunca del mismo. Igualmente, la suposición que dio paso a la contradicción de más arriba sobre la clase de todos

los objetos, consistente en permitir que las subclases de una clase sean también miembros suyos, queda rechazada: como los miembros de una clase deben pertenecer al mismo tipo, tales subclases no pueden acceder al estatuto de miembros; con ello no podríamos ya formar la nueva clase, compuesta por los objetos más todas las posibles subclases formadas por todas sus combinaciones posibles. En consecuencia, hablar del número cardinal máximo reclamará necesariamente la referencia a un tipo concreto, y el cardinal máximo de cada tipo quedará superado por el cardinal del tipo siguiente, sin que podamos nunca «sumar» tales cardinales entre sí.

Metodológicamente hablando, la idea principal aquí expuesta se parece a la solución propuesta por la teoría russelliana de las descripciones definidas, siendo la diferencia principal que las expresiones que no respeten la jerarquía de tipos se declaran carentes de sentido, mientras los enunciados conteniendo expresiones descriptivas sin referencia se declaraban falsos, al menos una vez transcritos según dispone la teoría (aunque podría decirse que antes de realizar esa transcripción carecen también de sentido). Efectivamente, la teoría de las descripciones se desembaraza de las expresiones descriptivas como expresiones mal formadas, que se caracterizan por dar la impresión errónea de designar algo, impresión que queda deshecha una vez que tales expresiones son convenientemente reescritas. Similarmente, la teoría de los tipos, al menos en cuanto «teoría sin clases» (pues las clases no pueden considerarse «objetos», a la par con los «individuos»), señala el sinsentido de considerar a los símbolos de clases como símbolos con referencia. Así, los enunciados que parecen tratar acerca de clases, al carecer de sentido, sólo alcanzarán significación una vez transcritos de forma que ya no contengan ningún símbolo supuestamente referido a una clase. Estamos, pues, ante un claro paralelismo metodológico, donde ciertos fragmentos del lenguaje ordinario deben reescribirse hasta reflejar la realidad con mayor transparencia, partiendo del necesario proceso de análisis conceptual, apoyado en técnicas lógicas. En la jerga de Russell, tanto los símbolos de clases como los descriptivos son símbolos «incompletos». En el marco de su programa filosófico general, heredero de Bradley y Frege, y heredado por el primer Wittgenstein: la forma gramatical/aparente ha de ser superada hasta hallar la forma lógica/real subyacente.

Junto a cualquier posible valoración de la teoría de los tipos desde la óptica técnica, conviene recordar que para Russell tal teoría pretendía también cumplir fines filosóficos, como puede verse cuando se reflexiona sobre el paralelismo metodológico ya señalado con la teoría de las descripciones. En efecto, la teoría de los tipos no sólo esta-

ba destinada a resolver las paradojas, sino que también debía satisfacer nuestro sentido común lógico, lo que para Russell significaba, más o menos, el coincidir con nuestras intuiciones preteóricas sobre ontología y semántica. Es por ello por lo que la teoría russelliana de los tipos, a diferencia, por ejemplo, del mecanismo de estratificación mediante subíndices propuesto por Quine, destinado *únicamente* a evitar las antinomias, debía presentar un marco donde la ontología atomística extraída del análisis filosófico del lenguaje, y constituido esencialmente por individuos (sujetos) y propiedades (predicados), pudiera hallar terreno abonado para desarrollarse armónicamente. Ése fue el papel del esquema, procedente de la distinción fregeana entre concepto y objeto, que lo divide todo en funciones proposicionales (propiedades) y sus argumentos (individuos), esquema que a la larga daría cuenta tanto de las descripciones como de las clases, y con ellas de toda la matemática, presentando, al mismo tiempo, una materia prima ideal para la inserción de la teoría de los tipos en el sistema de lógica matemática de los *Principia*.

Curiosamente, Russell solía insistir en que para el funcionamiento correcto de la teoría de los tipos se necesita sólo que la jerarquización por niveles sea correcta, pero no que la naturaleza de los objetos del primer nivel sea conocida, lo que constituiría una cuestión más de orden ontológico que lógico. Esto hay que tomarlo, creo yo, *cum grano salis*, pues podría llevar a interpretar la teoría como simple argucia útil para evitar las paradojas, y ello la convertiría en blanco fácil para la crítica. Así, es habitual criticar la estratificación de Quine como mero recurso *ad hoc*. Por el contrario, no debe olvidarse que la teoría de los tipos es también, y quizá sobre todo, una teoría ontosemántica que divide el mundo en categorías últimas y sobre ellas calca un patrón de jerarquías funcionales y proposicionales.

La teoría de los tipos es en realidad mucho más complicada, debido a la necesidad de resolver paradojas de un carácter distinto a las puramente lógicas: son las llamadas desde Ramsey paradojas semánticas, que llevaron al recurso de añadir una división adicional en órdenes, dando así lugar a nuevos problemas. De entre ellas, la más célebre es la que genera Epiménides el mentiroso al afirmar «Estoy mintiendo». Al hacer esa afirmación, como al hacer cualquier otra, o miente o dice la verdad. Ahora bien, si miente al decir que miente, entonces es que dice la verdad, pero, si al decir que miente dice la verdad, entonces es que, efectivamente, miente.

Se trata de una situación similar a la que genera la paradoja de las clases: en ambas nos encontramos con una entidad, en este caso una afirmación, que se engloba a sí misma; así como aceptar que una cla-

se pueda ser miembro de sí misma (autopertenencia) ocasiona la paradoja anterior, aceptar que el contenido de una afirmación pueda aplicarse a ella misma (autorreferencia) conduce directamente a la nueva paradoja, y de hecho a toda una extensa familia de ellas. Ese paralelismo llevó a Russell a hablar del «principio del círculo vicioso» (introducido por Poincaré) como origen último de todas las paradojas. Al formularlo Russell fue especialmente descuidado, pudiendo hablarse incluso de varias versiones del mismo, pero una formulación sencilla podría ser ésta: lo que presupone la totalidad de una colección no debe formar parte de esa colección. Ello da lugar a una solución estrictamente paralela a la ya descrita para la paradoja de las clases. Véamoslo.

Cuando Epiménides afirma que está mintiendo, lo que hace en realidad es decir que *todos* los enunciados que profiere son falsos, lo cual incluye obviamente a ese mismo enunciado. Pero la división de esa gran totalidad en distintos niveles, debidamente separados, evita fácilmente la autorreferencia. Así, cuando Epiménides desee decir que cierto enunciado es falso, o que los enunciados de cierto conjunto son falsos, debe siempre acotar ese enunciado, o ese conjunto de enunciados, de manera que lo dicho no se autoimplique. Por ejemplo, puede decir: «El enunciado p es falso»; llamemos q a este enunciado entrecomillado. A partir de ahí, si el enunciado p es de orden n, entonces q es de orden $n + 1$ y la paradoja no aparece. Igualmente, si Epiménides dice: «Todos los enunciados del conjunto s son falsos», está afirmando un enunciado que no está incluido en el conjunto s, sino que pertenece a un orden superior, con lo que no se da tampoco la paradoja.

La limitación del discurso a órdenes bien especificados tiene, sin embargo, una consecuencia indeseada para otro de los grandes objetivos de la teoría de los tipos, que era conservar lo máximo de la matemática al realizar su reducción a la lógica. No otra era la esencia del programa logicista russelliano: definir los conceptos matemáticos en términos de conceptos lógicos y demostrar las verdades de la matemática a partir de axiomas puramente lógicos. La consecuencia indeseada consiste en que a menudo conviene referirse a *todos* los objetos que cumplen cierta propiedad, o a *todas* las propiedades de cierto objeto (o a *todas* las funciones que cierto argumento hace verdaderas, o a *todas* las proposiciones relevantes a cierto concepto), con lo que parece que olvidamos la necesaria precaución de señalar a qué tipo o tipos nos estamos refiriendo.

En efecto, al definir los números al modo logicista, podríamos simplemente definir el número dos como la clase de *todos* los pares;

igualmente, al definir la identidad podríamos decir que dos objetos son idénticos cuando comparten *todas* las propiedades; por último, al introducir el principio del tercio excluso diríamos que *todas* las proposiciones son o verdaderas o falsas. Si, en nuestro deseo de respetar la teoría, tratásemos de limitar a ciertos tipos las afirmaciones así generalizadas, éstas perderían su interés, y obtendríamos un número dos distinto para cada tipo, una relación de idéntidad distinta para cada tipo y un principio del tercio excluso limitado al tipo u orden de las proposiciones consideradas, y ciertamente no era eso lo que el Russell logicista, que aspiraba a la máxima generalidad de la lógica y a la total universalidad de las variables consideradas, podía aceptar.

El remedio, que para algunos fue peor que la propia enfermedad, consistió en la introducción del llamado axioma de reducibilidad, o axioma de las clases (porque equivalía a admitir por la puerta trasera las clases, expulsadas hacía poco por la puerta principal mediante la teoría «sin clases»). Este axioma se introduce canónicamente para funciones proposicionales, que se definen como aquellas funciones que toman proposiciones como sus valores al darles un argumento. Semejantes a los conceptos de Frege y perfectamente insertables en el marco de la lógica de predicados heredada de Peano, las funciones proposicionales serían las entidades a las que finalmente se reduciría todo en el logicismo russelliano a título de materia prima. Lo que trata de hacer el axioma de reducibilidad es lograr un método para reducir el orden de cualquier función dada conservando la verdad de sus valores.

En concreto, se definen las funciones predicativas como aquellas que no envuelven ninguna referencia a una clase general de funciones o individuos (sus argumentos), sino sólo a las funciones o los individuos de un orden inmediato inferior. Así, una función predicativa de un individuo será una función de primer orden, y así sucesivamente, suponiendo el axioma de que existen funciones predicativas de cualquier orden. Pues bien, lo que afirma el axioma de reducibilidad es que toda función proposicional es siempre equivalente a alguna función predicativa para todos sus valores. Por tanto, podemos realizar tranquilamente generalizaciones sin miedo a las paradojas, pues siempre podremos señalar toda la pirámide de individuos o funciones implicados en cada ocasión, evitando así los problemas de autopertenencia o autorreferencia, y todo ello dentro de un mundo perfectamente ordenado, donde cada entidad guarda su sitio en la jerarquía y es siempre reducible a entidades de orden/tipo inferior.

El axioma de reducibilidad tenía claramente un aire limitativo, y por supuesto no podía presentarse como una verdad lógica, lo cual

era necesario para que el logicismo russelliano se desarrollara según lo deseado. Es incluso discutible si técnicamente funciona o no, pero como tanto su estatuto de supuesta verdad lógica como su funcionamiento son temas polémicos, y por tanto abiertos, los vamos a ignorar aquí. En cambio, es del mayor interés relacionar la limitación que introducía en el logicismo de Russell con otra limitación mucho más interesante para nosotros: la de la propia formulación de la teoría de los tipos. Repárese en que la formulación del axioma de reducibilidad es de un tipo de generalidad muy superior al de las entidades a las que se aplica, estableciendo que tales entidades están jerarquizadas y son reducibles verticalmente «hacia abajo». Para ello ha de adoptarse una óptica sumamente general, al estilo del «ojo de Dios», tantas veces rechazado por epistemólogos poco dados a admitir puntos de vista privilegiados.

Pues bien, en esa misma línea puede también decirse, incluso con mayor motivo, que la formulación de la propia teoría de los tipos pecaría contra sí misma. En efecto, para establecer la jerarquía de tipos, o la de órdenes, hemos necesariamente de referirnos a *todos* los objetos (que se hallan ordenados/jerarquizados en tipos: individuos, clases, etc.) y a *todas* las funciones (propiedades) y proposiciones, aunque sólo sea para señalar el hecho mismo de esa jerarquía; pero eso es precisamente lo que la teoría dice que es imposible sin caer en el sinsentido: así, la propia formulación pecaría contra sí misma.

Se trata, obviamente, del problema de la inefabilidad que tan hábilmente abordó Wittgenstein en el *Tractatus*. Para Wittgenstein las propias proposiciones del *Tractatus*, al no ser proposiciones sobre el mundo, constituirían estrictamente sinsentidos, aunque desde cierto punto de vista su sinsentido sería muy distinto del sinsentido de las proposiciones meramente absurdas: se trataría de proposiciones útiles para ser entendidas provisionalmente (la escalera que se arroja una vez alcanzada nuestra meta), a título de pistas para alcanzar una nueva visión general.

Pues bien, no por casualidad fue Wittgenstein quien primero formuló la crítica a los tipos según la cual éstos no pueden ni siquiera formularse, bajo pena de ser vulnerados *eo ipso*. Así, según Wittgenstein, aunque las limitaciones establecidas por la teoría son importantes, no pueden hacerse explícitas (no pueden *decirse*), sino que el propio simbolismo utilizado debe *mostrarlas* sin necesidad de referirse a ellas.

En ello Wittgenstein probablemente seguía la línea ya establecida por Frege cuando señaló claramente que los símbolos de funciones no pueden hacer de sujetos, a riesgo de perder su sentido (la célebre pa-

radoja del concepto caballo, que no sería un concepto sino un objeto), o cuando estableció toda una ontología de objetos (sujetos) y funciones (predicados) según la cual no podemos unir a través de un tercer símbolo los símbolos para tales entidades últimas, pues en tal caso caeríamos en la famosa paradoja de Bradley contra las relaciones: unir dos entidades mediante una relación reclamaría, a su vez, entidades adicionales que unieran la relación a las entidades originales. Para Frege tal problema se evita simplemente estableciendo que los objetos son entidades saturadas y las funciones entidades insaturadas, que como tales «encajan» entre sí a la perfección, sin necesidad de hacer explícita esa unión mediante una tercera entidad, como ya había avisado muchos siglos antes Aristóteles mediante su conocido argumento del tercer hombre.

El propio Russell, sin embargo, había intuido esa autolimitación de la propia teoría de los tipos, y sus dificultades para ser formulada, cuando ya en 1908 escribió:

> No podemos decir: «cuando hablo de todas las proposiciones, quiero decir todas excepto aquellas en que se alude a «todas las proposiciones»; pues en esta aclaración habríamos hecho mención de las proposiciones en que son mencionadas todas las proposiciones, cosa que no nos es posible llevar a cabo con sentido (Russell, 1908, 83).

La solución apuntada por Wittgenstein de que debe ser el propio simbolismo el que, de manera implícita, realice por sí mismo la distinción entre lo significativo y lo no significativo fue ya introducida claramente por Russell, también en 1908:

> Una variable no podría ser nunca restringida a cierto campo si la función proposicional en que interviene aquélla conservase su significación cuando la variable quede fuera de dicho campo. Pero si la función dejara de ser significativa tan pronto como la variable sobrepasase los límites de un cierto campo, la variable quedaría entonces *ipso facto* circunscrita a dicho campo sin necesidad de ninguna declaración explícita a esos efectos. Este principio ha de tenerse presente a lo largo del desarrollo de la teoría de los tipos (*ibid.*, 99).

Curiosamente, el tema de haber tratado de expresar lo inefable, a pesar de todas las precauciones explícitamente tomadas por Russell con su teoría de los tipos, sale de nuevo a relucir, en sentido inverso, cuando Russell criticó la propia teoría de la inexpresabilidad de lo místico (del mostrar el todo del mundo) del *Tractatus*. Para Wittgenstein, las proposiciones con sentido sólo pueden versar sobre ciertas parcelas del mundo, que son las que resultan figuradas por aquéllas.

Pero lo que hace posible que la proposición figure esa realidad, que es la forma lógica que comparte con ella, es inefable: al intentar hacerla explícita tratamos erróneamente de convertirla en otra realidad a ser figurada, con lo que se necesitaría un nuevo medio de expresión: una nueva forma lógica, y caeríamos en el regreso infinito *a la* Bradley-Frege. Del mismo modo, las proposiciones «metateóricas» del *Tractatus*, aunque carentes estrictamente de sentido, pues no reflejan estados de cosas posibles, se limitan a dar pistas indirectas sobre una correcta acomodación filosófica ante el mundo y el lenguaje; pero nada general puede ser dicho sobre el mundo y sobre el lenguaje en la línea estricta de la figuración: la semántica es inefable.

Para Russell esa supuesta inefabilidad es una especie de trampa, pues sugiere que, a pesar de todo, Wittgenstein se las arregló para decir una buena cantidad de cosas sobre lo que en puridad nada puede decirse. Pues bien, ése es el punto exacto donde Russell, sorprendentemente, parece constituir un claro precedente de la distinción tarskiana entre lenguajes, ya en la introducción a la edición inglesa del *Tractatus* en 1922:

> Estas dificultades me sugieren la siguiente posibilidad: que todo lenguaje tiene, como Wittgenstein dice, una estructura de la cual nada puede decirse *en el lenguaje*, pero que puede haber otro lenguaje que trate de la estructura del primer lenguaje y que tenga una nueva estructura y que esta jerarquía de lenguaje no tenga límites (p. 28).

Naturalmente, las observaciones de la máxima generalidad de Wittgenstein podrían entonces aplicarse a esa totalidad de lenguajes, al menos en cierto sentido más o menos místico. Russell halla, pues, una forma de aplicar el principio básico de la teoría de los tipos, el de la jerarquización ordenada de niveles, para superar algo que el propio Wittgenstein había esgrimido contra la propia teoría de los tipos: su propia incapacidad de ser formulada, y ello valiéndose hábilmente de una crítica propia a la confesada imposiblidad de un estatuto aceptable para las proposiciones más generales del *Tractatus*.

Esto nos lleva, dando un paso más en el orden de la generalidad, al problema de hasta qué punto la teoría de los tipos supone una limitación a lo que debía resultar completamente ilimitado en los *Principia*: la universalidad del ámbito de la variable. Puesto que la teoría de los tipos establece una jerarquización en ese ámbito de significación, a través de la jerarquía de las funciones proposicionales, ello parece pecar contra uno de los fundamentos de la misma esencia lógica de los *Principia*: la variable debe ser universal, totalmente irrestricta,

como reclama la concepción russelliana de la lógica, entendida no como mero cálculo, sino como lenguaje universal inescapable, por usar la ya clásica distinción establecida por van Heijenoort respecto a Frege, Russell y Wittgenstein.

En efecto, si las variables de las funciones proposicionales deben restringirse, para evitar las paradojas, a cierto ámbito de significación, entonces, cuando en nuestras expresiones aparezca la cuantificación sobre individuos, propiedades y proposiciones, ésta debe estar restringida convenientemente, de forma que el discurso sobre «todos» ellos resulte prohibido. Como hemos visto arriba, la solución de Russell consiste en restringir ese campo de significación de manera automática, es decir, sin ninguna declaración explícita al respecto (que vulneraría la teoría), sino haciendo que el uso mismo del simbolismo lo deje suficientemente claro.

La idea vuelve, pues, de nuevo sobre el tema de la inefabilidad: al leer el simbolismo para la variable, lo que veremos es «una limitación *interna* de x, facilitada por la propia naturaleza de la función; y se trata de una limitación que no requiere ser explícitamente formulada, ya que no cabe a una función ser verdadera con mayor generalidad que para todos sus valores» (1908, 98). Desgraciadamente Russell parece implementar tal limitación apoyándose en la distinción entre «todos» (prohibido) y «cualquier» (permitido), que resulta escasamente convincente.

Así, finalmente, el tema de la generalidad de nuestras afirmaciones dentro de la teoría, sobre la teoría y sobre sus objetos, se muestra como la dificultad quizá insuperable tanto para su funcionamiento como incluso para su simple formulación consistente. Curiosamente, los resultados metamatemáticos posteriores a los *Principia* atacaron los fundamentos de ese monumento a la idea logicista precisamente desde la misma óptica. Gödel demostró que un sistema no puede ser completo, no puede aspirar a reducir «todas» las proposiciones de la matemática a la lógica, pues en sistemas de semejante calado es siempre posible construir un enunciado sobre los números naturales que, siendo obviamente verdadero bajo la interpretación usual del simbolismo, no puede demostrarse mediante los axiomas y las reglas de inferencia del sistema. Asimismo, demostró que si los *Principia* son consistentes, si «todas» sus proposiciones son consistentes entre sí (si no hay contradicción entre ellas), tal consistencia no puede demostrarse dentro del sistema. Por tanto, aun si la teoría de los tipos hubiese funcionado técnicamente, y aun si sus fundamentos filosóficos hubiesen sido convincentes, estaba llamada a otro tipo de fracaso como destino final.

BIBLIOGRAFÍA

Copi, I. (1971), *The theory of logical types*, Kegan Paul, London.

Chihara, C. S. (1973), *Ontology and the vicious circle principle*, Ithaca, Cornell University Press, New York.

Garciadiego, A. (1992), *Bertrand Russell and the origins of the set-theoretic paradoxes*, Birkhäuser, Basel/Boston/Berlin.

Gödel, K. (1944), «Russell's mathematical logic», en Schilpp, 123-154.

Gödel, K. (1994), *Ensayos inéditos*, en F. Rodríguez Consuegra (ed.), Grijalbo-Mondadori, Barcelona.

Henkin, L. (1962), «Are logic and mathematics identical?»: *Science* 138, 788-794.

Hurtado, G. (1998), *Proposiciones russellianas*, UNAM, México.

Kneale, W. y Kneale, M. (1961), *The development of logic*, Clarendon Press, Oxford. Trad. cast. de J. Muguerza, Tecnos, Madrid, 1972.

Landini. G. (1998), *Russell's hidden substitutional theory*, OUP, New York/Oxford.

Rodríguez Consuegra, F. (1989), «Russell's theory of types, 1901-1910: its complex origins in the unpublished manuscripts»: *History and Philosophy of Logic* 10, 131-164.

Rodríguez Consuegra, F. (1991a, ^{2}1993), *The mathematical philosophy of Bertrand Russell: origins and development*, Birkhäuser, Basel/Boston/Berlin.

Rodríguez Consuegra, F. (1991b), «El logicismo russelliano: su significado filosófico»: *Crítica* 67, 15-39.

Rodríguez Consuegra, F. (1994), «Kurt Gödel y la filosofía de la matemática», en *Estudio introductorio a Gödel*, 22-125.

Rodríguez Consuegra, F. (1998), «Russell, Gödel and logicism», en A. Irvine (ed.), *Bertrand Russell: Critical Assessments*, 4 vols., vol. II, Routledge, London/New York, 320-329.

Rodríguez Consuegra, F. (1999), «Bradley, Frege and relatedness»: *Bradley Studies* 5, 113-125.

Russell, B. (1903), *The principles of mathemathics*, CUP, London. (Reimp., con nueva introducción: Allen & Unwin, London, 1937.)

Russell, B. (1908), «Mathematical logic as based on the theory of types»: *Amer. Jrn. Math.* 30, 222-62. (Reimp. en 1956, 59-102.)

Russell, B. (1910) (con A. N. Whitehead), *Principia mathematica*, vol. I, CUP, London. (Los vols. II y III en 1912 y 1913 respectivamente. 2.ª ed., con nueva introducción, de Russell solo, en 1927.)

Russell, B. (1918), «The philosophy of logical atomism»: *Monist* 28, 495-527; 29 (1919), 33-63, 190-222, 344-80. (Reimp. en 1956, 177-281.)

Russell, B. (1919), *Introduction to mathematical philosophy*, Allen & Unwin, London.

Russell, B. (1956), *Logic and knowledge*, en R. C. Marsh (ed.), Allen & Unwin, London. (Trad. cast. de J. Muguerza, Taurus, Madrid, 1966.)

Sainsbury, R. M. (1979), *Russell*, Routledge, London.

Schilpp, P. A. (ed.) (1944) ([4]1971), *The philosophy of Bertrand Russell*, Open Court, La Salle, Ill.
van Heijenoort, J. (1967), «Logic as language and logic as calculus»: *Synthese*, 17, 324-330.
Weiss, P. (1928), «The theory of types»: *Mind*, 37, 338-348.
Wittgenstein, L. (1922), *Tractatus Logico-Philosophicus*. (Trad. cast. de E. Tierno, Alianza, Madrid, 1973. Nueva y mejor trad. de J. Muñoz e I. Reguera, Alianza, Madrid, 1987.)
Wittgenstein, L. (1974), *Letters to Russell, Keynes and Moore*. (Trad. cast. de N. Míguez, Taurus, Madrid, 1979.)

LÓGICA Y TEORÍA DE CONJUNTOS

Jesús Mosterín

1. EL UNIVERSO CONJUNTISTA

A diferencia de la ciencia antigua, que buscaba una comprensión cualitativa de los fenómenos, la ciencia moderna se basa en la construcción de modelos teóricos (es decir, matemáticos) de la realidad. La realidad es excesivamente compleja para poder ser directamente comprendida por nuestras limitadas entendederas. Lo único que podemos hacer es buscar en el universo matemático una estructura que se parezca en algún aspecto relevante a la porción de la realidad que nos interesa, y usar esa estructura como modelo teórico simplificado de la realidad. Una vez que disponemos de un modelo teórico, podemos traducir al lenguaje de las matemáticas las preguntas que nos hacemos en la vida real, podemos computar la respuesta dentro del modelo y, finalmente, podemos retraducir esa respuesta matemática al lenguaje de la vida real.

Si queremos calcular trayectorias de aviones o barcos sobre la superficie terrestre, modelamos la Tierra mediante una esfera o un elipsoide de revolución. La forma real de la superficie terrestre es mucho más complicada; incluye cada hoja de cada árbol y cada espuma de cada ola; además, varía constantemente con cada brisa que sopla, con cada automóvil que pasa, con cada gota de lluvia que cae. Ninguna geometría conocida ni imaginable podría dar cuenta cabal de forma tan compleja. Ni falta que hace. Los modelos de la esfera o del elipsoide nos proporcionan aproximaciones suficientes para nuestras necesidades. En las teorías científicas avanzadas las estructuras matemáticas que utilizamos como modelos son bastante más complicadas. La cosmología usa la teoría general de la relatividad, que modela el es-

pacio-tiempo físico como una variedad diferencial provista de una cierta métrica (un campo tensorial). La mecánica cuántica modela los sistemas atómicos como espacios de Hilbert (espacios vectoriales complejos de un número infinito de dimensiones). ¿De dónde sacamos esas esferas y elipsoides, de dónde sacamos los números, los vectores, las probabilidades, las variedades diferenciales, los campos tensoriales, los espacios de Hilbert? Los sacamos del universo matemático, es decir, del universo conjuntista.

La ciencia teórica se aparta drásticamente de las formas habituales de comprensión. Para darse uno cuenta de cuán remoto es su método de representación, basta con considerar lo alejado que está el universo conjuntista (el depósito de todas las estructuras matemáticas) de los usos del lenguaje ordinario y de los hábitos de la experiencia cotidiana. Los matemáticos crean el universo conjuntista casi de la nada, a partir del conjunto vacío, mediante iteraciones sucesivas de ciertas operaciones.

Hacia 1925-1930 von Neumann y Zermelo introdujeron la jerarquía acumulativa como una manera transparente de visualizar el universo conjuntista, concebido como el resultado de un proceso inacabable que, partiendo del conjunto vacío, va generando nuevos conjuntos a partir de otros ya obtenidos anteriormente mediante la aplicación repetida de las operaciones de conjunto potencia y gran unión. La generación se produce paso a paso, siguiendo la serie de los ordinales. Los ordinales incluyen la sucesión de los números naturales y la extienden a lo transfinito. Los números ordinales son o bien el 0 o bien los ordinales sucesores (un ordinal sucesor es un ordinal β tal que, para algún ordinal α, $\beta = \alpha+1$) o bien los ordinales límite (un ordinal límite λ es un ordinal que no es el 0 ni es un ordinal sucesor). Von Neumann introdujo un método para definir funciones ordinales por recursión transfinita y lo aplicó en especial a la definición rigurosa de la jerarquía acumulativa de los conjuntos mediante la función ordinal V. Empezando por el conjunto vacío $\varnothing$, vamos construyendo paso a paso todos los conjuntos. En cada paso, indexado por el ordinal α, se obtiene un nuevo escalon, $V(\alpha)$.

$$V(0) = \varnothing$$
$$V(\alpha+1) = \wp V(\alpha)$$
$$V(\lambda) = \sup\{V(\beta): \beta < \lambda\} = \bigcup_{\beta < \lambda} V(\beta)$$

La gran unión de todos estos escalones, $\bigcup_{\alpha \in \Omega} V(\alpha)$, constituye la *jerarquía acumulativa*. Aquí Ω simboliza la clase de todos los números ordinales. El *axioma de regularidad* equivale a decir que cada con-

junto está en alguno de estos escalones, $\forall x \exists \alpha \; x \in V(\alpha)$, lo cual determina el rango del conjunto. También equivale a identificar el universo conjuntista V con la jerarquía acumulativa, $V = \bigcup_{\alpha \in \Omega} V(\alpha)$. Mientras Frege y Russell pensaban en los conjuntos básicamente como las extensiones de los conceptos, los lógicos y matemáticos actuales tienden a pensar en ellos más bien como los componentes de la jerarquía acumulativa.

El universo conjuntista es la clase de todos los conjuntos, que encierra en su seno todos los conjuntos imaginables e inimaginables pero posibles, todas las entidades, funciones, espacios, sistemas y estructuras matemáticas. En la teoría de conjuntos actual suele aceptarse el axioma de regularidad, lo que permite visualizar el universo conjuntista como un cono invertido, cuyo vértice es el conjunto vacío y cuyo eje está constituido por la serie de los ordinales. El cono va creciendo y ensanchándose hacia arriba mediante iteraciones sucesivas de las operaciones de conjunto potencia y gran unión, indexadas por los ordinales. A cada ordinal α corresponde una nueva rodaja de conjuntos, los conjuntos de rango α. La unión de todas estas rodajas forma la jerarquía acumulativa, identificada con el universo conjuntista.

Respecto al cono invertido del universo conjuntista podemos preguntarnos (1) cómo es de ancho y (2) cómo es de alto. Diversos axiomas e hipótesis determinan la anchura y la altura del cono. El axioma de constructibilidad ($V = L$) y la hipótesis del continuo ($2^{\aleph_0} = \aleph_1$) se refieren a la anchura del cono. Si aceptamos la hipótesis del continuo, el cono se estrecha. Si aceptamos el axioma de constructibilidad, se estrecha todavía más. Si rechazamos el axioma de constructibilidad, el cono se ensancha; si rechazamos la hipótesis del continuo, el cono se ensancha aún más. Podría haber sólo ordinales finitos, en cuyo caso el cono tendría poca altura. El axioma de infinitud lo hace más alto. Además, una vez que tenemos ω, la serie de los ordinales transfinitos crece ilimitadamente. ω, $\omega+1$, ... $\omega+\omega$, ... tienen la cardinalidad infinita mínima, $\aleph_0$. Pero $\wp\omega$ tiene ya cardinalidad $\aleph_1$ o superior. Los axiomas habituales de la teoría de conjuntos no determinan hasta dónde y hasta qué cardinalidades lleguen los ordinales y, con ellos, el universo conjuntista. No determinan si existen los cardinales monstruosamente grandes que podemos definir. Para ello hacen falta axiomas específicos que postulen su existencia. Los axiomas de cardinales grandes se refieren a la altura del cono. La existencia de cardinales inaccesibles, de cardinales de Mahlo, de cardinales débilmente compactos, de cardinales medibles, de cardinales de Woodin o de cardinales supercompactos dispara la altura del cono a extremos de vértigo.

La teoría de conjuntos está llena de cuestiones abiertas. Según la respuesta que demos a esas preguntas, el árbol teórico se ramifica en una frondosa variedad de teorías alternativas de conjuntos. Afortunadamente, todas esas teorías alternativas poseen una intersección común, que basta para las necesidades de la matemática habitual, es decir, de la matemática que tiene aplicaciones extramatemáticas. Sin embargo, si profundizamos suficientemente en el mundo de los conjuntos, enseguida topamos con preguntas cuya respuesta cae fuera de esa intersección teórica común. Las posibles respuestas alternativas son independientes de los axiomas habituales (como ZFC o NBG). En comparación con el océano proceloso de la teoría de conjuntos, la lógica parece algo inofensivo y trivial, al margen de los supuestos y discusiones conjuntistas; pero no es así, como veremos a continuación.

2. LÓGICA DE SEGUNDO ORDEN

En la lógica de orden cero (la lógica proposicional) no hay cuantificación alguna. En la de primer orden podemos cuantificar sobre individuos, pero no sobre conjuntos de individuos (o relaciones entre individuos). En la lógica de segundo orden, finalmente, podemos cuantificar tanto sobre los individuos del universo del discurso como sobre los subconjuntos de ese universo, es decir, sobre los conjuntos de individuos y las relaciones entre individuos. La lógica de primer orden posee un solo tipo de variables, que varían sobre (o se refieren indistintamente a) los individuos del universo de la interpretación. Estas variables individuales pueden ser cuantificadas por los cuantificadores. La lógica de primer orden acomoda también la presencia de relatores o parámetros que, en una interpretación dada, se refieren a ciertas propiedades o relaciones determinadas, pero carece de recursos para referirse a subconjuntos cualesquiera del universo de la interpretación. Esos recursos son característicos de la lógica de segundo orden, que incluye variables que varían sobre subconjuntos o propiedades cualesquiera y permite cuantificarlas. El enunciado «Napoleón tenía todas las cualidades de un buen general» no puede traducirse al lenguaje de la lógica de primer orden, pero sí al de segundo orden: $\forall Z\ (\forall x\ (Bx \Rightarrow Zx) \Rightarrow Zn)$, donde B significa ser un buen general y n se refiere a Napoleón. En este ejemplo, Z es una variable de segundo orden, que representa propiedades cualesquiera. El principio de inducción aritmética dice que todos los números naturales poseen cualquier propiedad poseída por 0 y tal que, siempre que la tenga un número cualquiera x, también la tiene su siguiente, $s(x) = x+1$. Este

principio se traduce naturalmente al lenguaje de segundo orden: $\forall Z$ $(Z0 \wedge \forall x\ (Zx \Rightarrow Zs(x)) \Rightarrow \forall x\ Zx)$, donde las variables individuales (minúsculas) varían sobre los números naturales, mientras que la variable de segundo orden Z varía sobre conjuntos cualesquiera de números naturales.

Hay diversos tipos de lenguajes formales de segundo orden. Por ejemplo, unos pueden tener tanto variables predicativas como functoriales, mientras otros se limitan a las variables predicativas, o incluso a las variables predicativas unarias. Aquí consideraremos lenguajes formales con variables predicativas *n*-arias cualesquiera y sin variables functoriales. El alfabeto consta de las mismas constantes lógicas y eventuales parámetros que la lógica de primer orden, pero posee, además de las variables individuales, variables predicativas *n*-arias para cada número *n*. Entre las constantes lógicas primitivas se puede prescindir del signo de identidad, ya que es definible en la lógica de segundo orden. Los *términos* se definen como en la lógica de primer orden. Una *fórmula* de la lógica de segundo orden es una secuencia finita de signos del alfabeto de la lógica de segundo orden construida de acuerdo con las siguiente reglas: 1) Si R es un relator (un signo de relación) *n*-ario y $\tau_1...\tau_n$ son términos, entonces $R\tau_1...\tau_n$ es una fórmula. 2) Si W es una variable *n*-aria y $\tau_1...\tau_n$ son términos, entonces $W\tau_1...\tau_n$ es una fórmula. 3) Si φ es una fórmula, entonces $\neg\varphi$ también es una fórmula. 4) Si α y β son fórmulas, entonces $(\alpha \wedge \beta)$, $(\alpha \vee \beta)$, $(\alpha \Rightarrow \beta)$ y $(\alpha \Leftrightarrow \beta)$ son también fórmulas. 5) Si φ es una fórmula y x es una variable individual, entonces $\forall x\ \varphi$ y $\exists x\ \varphi$ también son fórmulas. 6) Si φ es una fórmula y W es una variable predicativa (de cualquier número ario), entonces $\forall W\ \varphi$ y $\exists W\ \varphi$ también son fórmulas. Si la última regla usada en la construcción de una fórmula fue la regla (6), hablamos de una cuantificación (universal o existencial) de segundo orden. Las variables, tanto las individuales como las predicativas, pueden estar libres o ligadas en las fórmulas. Una fórmula sin variables libres es una sentencia.

Una fórmula pura de la lógica de segundo orden es una fórmula de segundo orden sin parámetros, es decir, escrita sólo con variables, constantes lógicas y paréntesis. Una sentencia pura de la lógica de segundo orden carece de símbolos específicamente matemáticos o conjuntistas (como $\in$), sólo posee constantes lógicas y variables cuantificadas. Por tanto, la validez lógica de una sentencia pura de segundo orden parece un asunto bastante apartado de la teoría de conjuntos, aunque, como veremos, no es así.

La semántica de la lógica de segundo orden es como la de primer orden, excepto en lo que respecta a las variables predicativas, que se

interpretan como variando sobre subconjuntos cualesquiera del universo A de la interpretación (es decir, sobre $\wp A$) en el caso de las variables predicativas unarias y, en el caso general de las n-arias, como variando sobre subconjuntos cualesquiera del producto cartesiano, iterado n-1 veces, del universo consigo mismo (es decir, sobre $\wp A^n = \wp(A \times ... \times A)$, donde A aparece n veces).

Sea $\mathcal{L}$ un lenguaje de segundo orden con tipo de semejanza σ (donde σ es una secuencia de números que indica el número, tipo y aridad de los parámetros de $\mathcal{L}$) y sea A un sistema con universo A y homólogo a $\mathcal{L}$, es decir, con el mismo tipo de semejanza σ. Una interpretación $\mathfrak{I}$ sobre A asigna individuos de A a las variables individuales de $\mathcal{L}$ y relaciones sobre A a las variables predicativas de $\mathcal{L}$ y entidades distinguidas de A a los parámetros de $\mathcal{L}$. Sea g: {Variables individuales} $\rightarrow A$ una asignación de individuos de A a las variables individuales. Para cada n, sea h_n: {Variables predicativas n-arias} $\rightarrow A^n$ una asignación de relaciones n-arias sobre A a las variables predicativas. Sean c_j los nombres de $\mathcal{L}$, sean R_i los relatores de $\mathcal{L}$ y sean f_j los functores de $\mathcal{L}$. Sea $\mathsf{A} = \langle A, ... \mathfrak{I}(c_j)..., ... \mathfrak{I}(R_i)..., ... \mathfrak{I}(f_j)...\rangle$, donde para cada nombre c_j, $\mathfrak{I}(c_j) \in A$, para cada relator n-ario R_i, $\mathfrak{I}(R_i) \in \wp(A^n)$ y para cada functor n-ario f_j, $\mathfrak{I}(f_j)$: $A^m \rightarrow A$. Entonces, $\mathfrak{I} = \langle g, \{h_i\}, \mathsf{A}\rangle$ es una interpretación de $\mathcal{L}$ sobre A.

La interpretación $\mathfrak{I}$ asigna un individuo a cada término del lenguaje $\mathcal{L}$: $\mathfrak{I}(x) = g(x)$ para cada variable individual x. $\mathfrak{I}(f\tau_1 ... \tau_n) = \mathfrak{I}(f)\mathfrak{I}(\tau_1)...\mathfrak{I}(\tau_n)$. Para cada fórmula φ del lenguaje $\mathcal{L}$ está determinado (con la salvedad más abajo indicada) si la interpretación $\mathfrak{I}$ satisface φ o no satisface φ: 1) $\mathfrak{I}$ satisface $R\tau_1 ... \tau_n$ si y sólo si $(\mathfrak{I}(\tau_1) ... \mathfrak{I}(\tau_n)) \in \mathfrak{I}(R)$. 2) $\mathfrak{I}$ satisface $W\tau_1 ... \tau_n$ si y sólo si $(\mathfrak{I}(\tau_1) ... \mathfrak{I}(\tau_n)) \in h_n(W)$. 3) $\mathfrak{I}$ satisface $\neg\varphi$ si y sólo si $\mathfrak{I}$ no satisaface φ. 4) $\mathfrak{I}$ satisface $(\alpha \wedge \beta)$ si y sólo si $\mathfrak{I}$ satisface α y $\mathfrak{I}$ satisface β. 5) $\mathfrak{I}$ satisface $(\alpha \vee \beta)$ si y sólo si $\mathfrak{I}$ satisface α o $\mathfrak{I}$ satisface β. 6) $\mathfrak{I}$ satisface $(\alpha \Rightarrow \beta)$ si y sólo si $\mathfrak{I}$ satisface β o $\mathfrak{I}$ no satisface α. 7) $\mathfrak{I}$ satisface $(\alpha \Leftrightarrow \beta)$ si y sólo si ($\mathfrak{I}$ satisface α y $\mathfrak{I}$ satisface β) o ($\mathfrak{I}$ no satisface α y $\mathfrak{I}$ no satisface β). 8) $\mathfrak{I}$ satisface $\forall x\, \varphi$ si y sólo si para cada $a \in A$, la interpretación $\mathfrak{I}^{x:a} = \langle g', \{h_i\}, \mathsf{A}\rangle$ satisface a φ, donde g' asigna el individuo $a \in A$ a la variable x y coincide con g en todas sus otras asignaciones. 9) $\mathfrak{I}$ satisface $\exists x\varphi$ si y sólo si para algún $a \in$ A, $\mathfrak{I}^{x:a}$ satisface a φ. 10) Para cada variable predicativa n-aria W: $\mathfrak{I}$ satisface $\forall W\, \varphi$ si y sólo si para cada relación $R \in A^n$, la interpretación $\mathfrak{I}^{W:R} = \langle g, \{h_i : i \neq n\} \cup \{h_n'\}, \mathsf{A}\rangle$ satisface a φ, donde h_n' asigna la relación $R \in A^n$ a la variable W y coincide con h_n en todas sus otras asignaciones; y 11) $\mathfrak{I}$ satisface $\exists W\, \varphi$ si y sólo si para alguna relación $R \in A^n$, la interpretación $\mathfrak{I}^{W:R}$ satisface φ.

Sea Γ un conjunto de fórmulas de $\mathfrak{L}$. $\mathfrak{I}$ satisface Γ si y sólo si, para cada $\varphi \in \Gamma$, $\mathfrak{I}$ satisface φ. Sea Σ un conjunto de sentencias de $\mathfrak{L}$. Si la interpretación $\mathfrak{I}$ sobre A satisface Σ, entonces cualquier interpretación sobre A satisface a Σ. Por tanto, la satisfacción de una conjunto de sentencias sólo depende del sistema sobre el que la interpretamos, no de la interpretación particular (es decir, no depende de su concreta asignación de individuos a las variables individuales y de relaciones a las variables predicativas). Por ello podemos decir que el sistema A verifica o satisface Σ. Si el sistema A satisface Σ, decimos que A es una *realización* de Σ.

Las nociones semánticas de validez, satisfacibilidad, implicación, consecuencia e independencia se definen como en la lógica de primer orden. En especial una fórmula φ es *válida* si y sólo si es satisfecha por todas las interpretaciones; es *satisfacible* si es satisfecha por alguna interpretación; y es una *consecuencia* de un conjunto Γ de fórmulas si y sólo si cada interpretación que verifica Γ, verifica también φ.

La lógica de segundo orden permite definir muchas nociones indefinibles en la de primer orden, empezando por la identidad. Podemos introducir '=' como un signo definido mediante la definición $x = y \Leftrightarrow_{df} \forall Z\ (Zx \Leftrightarrow Zy)$. Incluso es posible definir la identidad de propiedades o conjuntos: $Z = W \Leftrightarrow_{df} \forall x(Zx \Leftrightarrow Wx$. En realidad, en la lógica de segundo orden se pueden definir muchas nociones típicamente conjuntistas, como la de que Z es el conjunto potencia de Y, abreviada Pot(Y,Z), o la de que un conjunto Y es biyectable con Z, abreviada $Y \sim Z$, o la de que un conjunto X es infinito, abreviada Inf(X). Por ejemplo, la biyectabilidad entre Y y Z se define así:

$$Y \sim Z \Leftrightarrow_{df} \exists W\ [\forall xyz\ ((Wxy \wedge Wxz \Rightarrow y=z) \wedge (Wyx \wedge Wzx \Rightarrow y = z) \wedge \forall u\ (Yu \Rightarrow \exists x\ (Wux \wedge Zx)) \wedge \forall x\ (Zx \Rightarrow \exists u\ (Wxu \wedge Yu))]$$

La infinitud de Z, siguiendo a Dedekind, se define así:

$$\mathrm{Inf}(Z) \Leftrightarrow_{df} \exists Y\ [\forall x\ (Yx \Rightarrow Zx) \wedge \exists x\ (Zx \wedge \neg Yx) \wedge Z \sim Y]$$

Que el conjunto potencia de X es Y, es decir, que $\wp(X) = Y$, simbolizado pot(X,Y), se define así:

$$\mathrm{pot}(X,Y) \Leftrightarrow_{df} \exists W\ [\forall u\ (Xu \Leftrightarrow \exists z\ Wuz) \wedge \forall u\ (Yu \Leftrightarrow \exists z\ Wzu) \wedge \forall Z\ (\forall u\ (Zu \Rightarrow Xu) \Rightarrow \exists v \forall x\ (Zx \Leftrightarrow Wxv)) \wedge \forall uv\ (\forall x\ (Wxu \Leftrightarrow Wxv) \Rightarrow u=v)]$$

Obsérvese que todas estas nociones son definibles mediante fórmulas puras de la lógica de segundo orden.

La lógica de segundo orden tiene una enorme potencia expresiva y permite caracterizar unívocamente (hasta isomorfía) las estructuras fundamentales de la matemática clásica, tales como el sistema de los números naturales, o el cuerpo ordenado de los números reales, o el espacio euclideo. Sea *s* un parámetro de operación unaria («el siguiente de»). El sistema estándar de los números naturales puede caracterizarse mediante estos tres axiomas:

$$\forall x\ s(x) \neq 0$$
$$\forall xy\ (s(x) = s(y) \Rightarrow x = y)$$
$$\forall Z\ (Z0 \wedge \forall x\ (Zx \Rightarrow Zs(x)) \Rightarrow \forall x\ Zx)$$

La clase de todas las consecuencias (en la lógica de segundo orden) de estos tres axiomas constituye la aritmética de Peano de segundo orden, que es una teoría categórica. Todas sus realizaciones son isomorfas entre sí. Por tanto, la aritmética de Peano es una teoría sintácticamente completa, que da respuesta a todas las preguntas aritméticas. A primera vista, parece como si la lógica de segundo orden fuera el marco formal adecuado para la aritmética, el análisis, la geometría euclidea e incluso la teoría de conjuntos, pues todas estas teorías, formalizadas en segundo orden, son completas y categóricas y caracterizan unívocamente sus respectivas estructuras. Sin embargo, ya el teorema de incompletud de Gödel nos advierte de que estamos ante un espejismo. Si la aritmética de segundo orden es completa y consistente, entonces no puede ser axiomatizable en el sentido técnico de recursivamente enumerable. Podemos, sí, escribir sus axiomas de segundo orden (acabamos de hacerlo), pero no hay manera de obtener todas las consecuencias de estos axiomas. Aunque dé respuesta a todas las preguntas aritméticas, no hay manera de encontrar esas respuestas de un modo sistemático. En efecto, no sólo no hemos encontrado un cálculo deductivo semánticamente completo para la lógica de segundo orden, sino que hemos demostrado que no puede encontrarse, pues no lo hay.

La validez lógica de segundo orden no es recursivamente enumerable, por lo que no hay manera efectiva de caracterizarla. No puede existir un cálculo deductivo completo para la lógica de segundo orden, por lo que es difícil sacarle jugo sintáctico. Tampoco valen para ella los teoremas de Löwenheim-Skolem y de compacidad, por lo que no da lugar a una semántica o teoría de modelos vigorosa, como la de primer orden. Además, y como habíamos insinuado al inicio, ni siquiera las nociones semánticas están bien definidas o determinadas, debido a la inextricable conexión de la lógica de segundo orden con

la teoría de conjuntos, que traslada a la primera todas las incertidumbres de la segunda. Por tanto, la noción misma de lógica de segundo orden no está extensionalmente bien determinada. La potencia expresiva manifestada en su capacidad de caracterizar categóricamente estructuras interesantes de la matemática clásica es una riqueza imposible de explotar en la práctica. Si supiéramos qué sentencias puras de la lógica de segundo orden son lógicamente válidas, podríamos zanjar muchas cuestiones abiertas de la teoría de conjuntos. Pero no hay manera de saberlo. Más bien al contrario, necesitaríamos zanjar previamente todas las cuestiones abiertas de la teoría de conjuntos para poder disponer de una lógica de segundo orden extensionalmente bien determinada.

3. AXIOMA DE ELECCIÓN

Para la mayor parte de las cuestiones abiertas y sin respuesta de la teoría de conjuntos hay una fórmula pura de la lógica de segundo orden, tal que esa fórmula es lógicamente válida si y sólo si la correspondiente cuestión conjuntista tiene una respuesta positiva. Nadie duda seriamente de los axiomas conjuntistas de ZF. Las primeras dudas graves se refieren al axioma de elección (AC), conjeturado por Cantor en 1883. El axioma de elección equivale al principio del buen orden, que era una pieza clave en la construcción cantoriana de la teoría de conjuntos, pues sólo si cada conjunto puede ser bien ordenado podemos estar seguros de que cada conjunto infinito tiene un ordinal como su tipo de orden y un alef como su cardinalidad. Aunque usado implícitamente con anterioridad, el axioma de elección fue introducido explícitamente por Ernst Zermelo en 1904 para probar el teorema del buen orden. Zermelo lo consideraba un principio lógico evidente y lo incluyó como axioma en su primera axiomatización de la teoría de conjuntos en 1908.

Sea A una familia (es decir, un conjunto) de conjuntos. Una función de elección sobre A es una función f que a cada conjunto no vacío $X \in A$ le asigna un elemento de X, $f(X) \in X$. Esta función elige o escoge un elemento de cada miembro no vacío de A. En el universo conjuntista cada conjunto (excepto el vacío) es una familia de conjuntos. El *axioma de elección* dice que para cada conjunto hay una función de elección. Formalmente,

$$\forall y \exists f\ (f \text{ es una función} \land \text{dominio}(f) = y \land \forall x\ (x \in y \land x \neq \varnothing \Rightarrow f(x) \in x))$$

A pesar de su carácter abstruso, el axioma de elección ha provocado intensas polémicas y despertado apasionados entusiasmos y aversiones entre los matemáticos. Ya desde su propuesta por Zermelo, tropezó con la oposición de Peano, Borel, Baire y Lebesgue. Los intuicionistas lo rechazaron, pues este axioma afirma la existencia de funciones de elección, sin darnos pista alguna de cómo construirlas. Algunos llegaron a temer que el axioma de elección pudiera introducir contradicciones en la teoría de conjuntos. Hoy sabemos que ese temor era infundado. En 1938 Gödel probó que el axioma de elección (AC) es compatible con los otros axiomas (por ejemplo, ZF) de la teoría de conjuntos, de tal modo que si ZF es consistente, también lo es ZF+AC (abreviado ZFC). Por tanto, el axioma de elección no introduce contradicciones que no estuviesen ya previamente presentes en la teoría de conjuntos sin axioma de elección. En 1963 Paul Cohen probó que la negación del axioma de elección (¬AC) también es compatible con el resto de los axiomas. Entre los dos, Gödel y Cohen demostraron que el axioma de elección es independiente de los otros axiomas. Por tanto, podemos aceptarlo o rechazarlo sin peligro de caer en contradicciones. Si actualmente la mayoría de los matemáticos lo aceptan, ello se debe a su notable fecundidad. No sólo es imprescindible para desarrollar la teoría de los cardinales transfinitos, sino que encuentra también múltiples aplicaciones en el álgebra, la topología y la teoría de la medida.

El axioma de elección (AC), en una de sus versiones, dice que existe una función universal de elección, es decir, una función que a cada conjunto no vacío le asigna uno de sus miembros. Esto equivale a decir que el universo de la interpretación puede ser bien ordenado. La siguiente sentencia pura de la lógica de segundo orden, interpretada sobre cualquier universo, dice que ese universo puede ser bien ordenado.

$$\exists W\,[\forall xyu\,(Wxy \wedge Wyu \Rightarrow Wxu) \wedge \forall x\,\neg Wxx \wedge \forall xy\,(Wxy \vee Wyx \vee x = y) \wedge \forall Z\,(\exists x\,Zx \Rightarrow \exists u\,(Zu \wedge \forall x\,(Zx \Rightarrow Wux \vee u = x)))]$$

Esta fórmula es satisfecha o verificada por una interpretación $\mathfrak{I}$ sobre un universo A si y sólo si A puede ser bien ordenado, es decir, si hay alguna relación que lo bien-ordena.

4. HIPÓTESIS DEL CONTINUO

Otra famosa y controvertida conjetura de Cantor es la hipótesis del continuo. En su versión generalizada (GCH), dice que para cada con-

junto infinito M, no hay una cardinalidad intermedia entre la de M y la del conjunto potencia de M, es decir, no hay un conjunto mayor que $|M|$ y menor que $|\wp(M)|$. Ya vimos antes que una cierta fórmula pura de segundo orden sirve para definir la infinitud, abreviada Inf, y otra sirve para definir el conjunto potencia, $\wp$. De modo similar se definen mediante fórmulas puras de segundo orden la menor o igual cardinalidad (es decir, la inyectabilidad), abreviada $\leq$, y la menor cardinalidad, abreviada $<$. GCH es equivalente a la siguiente fórmula pura de segundo orden, escrita aquí usando las abreviaturas Inf, pot, $<$ y $\leq$:

$$\forall XYZ\ (\mathrm{Inf}(X) \wedge \mathrm{pot}(X,Y) \wedge Z < Y \Rightarrow Z \leq X)$$

En la teoría de conjuntos ZF no podemos probar ni refutar AC o GCH, pues ambas hipótesis son independientes de los axiomas. Tanto AC como GCH son verdades lógicas (fórmulas válidas) de segundo orden, si son verdades conjuntistas. Pero no sabemos si lo son, o si queremos que lo sean, o si tiene sentido preguntarse si lo son.

5. CARDINALES INACCESIBLES

También la compleja noción conjuntista de cardinal inaccesible es definible mediante una fórmula pura de segundo orden.

Un cardinal κ es inaccessible si y sólo si

(1) κ es un cardinal innumerable (es decir, $\kappa > \aleph_0$),

(2) κ es un cardinal límite fuerte (es decir, la construcción del conjunto potencia (o la potencia de 2) de inferiores a κ no conduce a κ: para cada cardinal λ: $\lambda < \kappa$ implica $2^\lambda < \kappa$),

(3) κ es regular (es decir, κ no es el supremo de un conjunto de menos de κ ordinales menores que κ).

La regularidad de cardinales puede también ser definida en términos de cofinalidad. La cofinalidad de α, $\mathrm{cf}(\alpha)$, es el mínimo ordinal β tal que hay una función f: $\alpha \rightarrow \beta$ con recorrido de f cofinal en α. La cofinalidad de un ordinal límite λ, $\mathrm{cf}(\lambda)$, es el mínimo cardinal κ tal que λ es el supremo de κ cardinales menores. Un cardinal κ es regular si y sólo si $\mathrm{cf}(\kappa) = \kappa$.

Si κ es un cardinal inaccesible, entonces κ es un punto fijo de la función álef: $\aleph_\kappa = \kappa$. Si κ es un cardinal inaccesssible, entonces todos los axiomas de ZFC son verdad en V_κ (es decir, V_κ es un modelo de ZFC). Una consecuencia de este último hecho (y del segundo teorema de incompletud de Gödel) es que la existencia de cardinales inaccesibles es indemostrable en ZFC. La consistencia relativa también es

indemostrable. En ZFC no podemos demostrar ni refutar la existencia de cardinales inaccesibles. Por tanto, la aserción de que hay cardinales inaccesibles añade nueva fuerza a la teoría, es un axioma que añade nueva profundidad (o altura) al universo conjuntista.

Hay cardinales inaccesibles si y sólo si hay conjuntos de cardinalidad inaccesible. Abreviemos como Inacc(Z) el predicado que dice que el conjunto Z es de cardinalidad inaccesible. Este predicado puede ser definido en la lógica pura de segundo orden. (Recordemos que los símbolos Inf y $\wp$ son meras abreviaturas.)

$$\text{Inacc}(Z) \Leftrightarrow_{df} \text{Inf}(Z) \wedge \exists Y\,(Y < Z \wedge \text{Inf}(Y)) \wedge \forall X\,(X < Z \Rightarrow \exists Y\,(\text{pot}(X,Y) \wedge Y < Z) \wedge \forall XYW\,[X < Z \wedge \forall u\,(Xu \Leftrightarrow \exists x\,Wux) \wedge \forall u\,(Yu \Leftrightarrow \exists x\,Wxu) \wedge \forall y\,(Xy \Rightarrow \forall V\,(\forall x\,(Vx \Leftrightarrow Wyx) \Rightarrow V < Z)) \Rightarrow Y < Z]$$

Esta fórmula dice que Z es infinito e innumerable, que la construcción del conjunto potencia de inferiores a Z no nos conduce a Z (límite fuerte), y que el recorrido de cualquier función que aplica inferiores a Z a los elementos de inferiores a Z no nos conduce a Z (lo que equivale a la regularidad). Por tanto, la fórmula define un conjunto de cardinalidad inaccessible.

La fórmula pura de segundo orden $\exists X$ Inacc(X) es una fórmula satisfacible de la lógica de segundo orden si y sólo si hay cardinales inaccesibles. La fórmula pura de segundo orden $\neg\exists X$ Inacc(X) es una fórmula lógicamente válida si no hay cardinales inaccesibles. Pero no sabemos, ni podemos decidir a partir de los otros axiomas, si hay cardinales inaccesibles o no.

6. TEORÍA DE CONJUNTOS EN DISFRAZ

Diversos teóricos han propuesto otros axiomas todavía más potentes que el de los cardinales inaccesibles, que afirman la existencia de cardinales más y más grandes (como los cardinales de Mahlo, los cardinales débilmente compactos, los cardinales medibles, los cardinales de Woodin o los cardinales supercompactos). Cada uno de estos axiomas es independiente de los anteriores e implica a los anteriores. Con suficiente espacio y habilidad, podríamos reformular todos estos axiomas como sentencias de la lógica pura de segundo orden, cuya validez lógica depende de arcanos detalles de extensiones alternativas y controvertidas de la teoría de conjuntos habitual. Por tanto, la lógica de segundo orden puede considerarse como teoría de conjuntos disfrazada de lógica.

Podemos introducir un cálculo deductivo (semánticamente incompleto, pero correcto) para la lógica de segundo orden. Basta con agregar a un cálculo deductivo de la lógica de primer orden algunas nuevas reglas de inferencia de segundo orden que preserven la validez. Por ejemplo, podemos añadir reglas de inferencia como la que autoriza a pasar de $\forall W\, \varphi$ a $\varphi(R)$, donde W es una variable predicativa n-aria, R es un signo de relación n-aria y $\varphi(R)$ es el resultado de sustituir todas las apariciones libres de W en φ por R, o la que permite pasar de $\varphi(R)$ a $\exists W\, \varphi$. El cálculo resultante nos permite generar fórmulas válidas, aunque no todas, y deducir consecuencias, aunque no todas, a partir de premisas dadas. Este cálculo no es semánticamente completo, al menos en el sentido habitual, basado en las interpretaciones normales o estándar; sin embargo, puede ser semánticamente completo respecto a otra semántica alternativa, introducida por Henkin en 1950 y basada en las interpretaciones generales.

En una *interpretación general* (en el sentido de Henkin) del lenguaje $\mathcal{L}$ de segundo orden las variables predicativas unarias ya no se interpretan como variando sobre subconjuntos cualesquiera del universo A de la interpretación, es decir, sobre el conjunto potencia $\wp A$ entero, sino sólo sobre una porción de $\wp A$. Y lo mismo ocurre con las demás variables predicativas. Una relación n-aria R es definible en $\mathcal{L}$ si y sólo si existe una fórmula abierta φ de $\mathcal{L}$ con n variables libres tal que R es el conjunto de los n-tuplos de elementos de A que satisfacen esa fórmula φ. Una interpretación general fija arbitrariamente, para cada n, el conjunto $A_n \subseteq \wp A^n$ de las relaciones n-arias permisibles en esa interpretación, con la única condición de que las relaciones definibles estén entre las permisibles. Una interpretación estándar es un caso extremo de interpretación general, aquel en que todas las relaciones posibles son permisibles, es decir, donde $A_n = \wp A^n$ para cada n. Todas las interpretaciones estándar son generales, pero no a la inversa. Una fórmula es válida en sentido general si y sólo si es satisfecha por todas las interpretaciones generales. Hay más interpretaciones generales que interpretaciones estándar. Por consiguiente, hay menos fórmulas satisfechas por todas las interpretaciones generales que por todas las interpretaciones estándar, es decir, hay menos fórmulas válidas en sentido general que en sentido estándar, muchas menos. Tanto es así, que el conjunto de las fórmulas válidas en sentido general es lo suficientemente pequeño como para ser recursivamente enumerable. Por eso podemos ofrecer un cálculo deductivo semánticamente completo en sentido general que lo genera. Además, la lógica de segundo orden, interpretada en el sentido general de Henkin, satisface los teoremas de compacidad y de Löwenheim-Skolem. En

cierto modo, es lógica de primer orden en disfraz. Por tanto, todas las ventajas semánticas que prometía la semántica estándar de segundo orden desaparecen. En el lenguaje formal de segundo orden, interpretado en el sentido general, ya no se pueden caracterizar unívocamente las estructuras de la matemática clásica ni se pueden formalizar sus teorías de un modo categórico y completo.

7. LÓGICA DE PRIMER ORDEN Y CONJUNTOS INFINITOS

Es difícil sustraerse a la conclusión de que la lógica de segundo orden es teoría de conjuntos en disfraz. Sin embargo, frecuentemente se piensa que la lógica de primer orden no está contaminada por consideración conjuntista alguna. ¿Hasta qué punto es así?

El conjunto de las sentencias válidas de la lógica de primer orden es el conjunto de todas las sentencias de primer orden que son satisfechas (o verificadas o verdad) en todos los sistemas o estructuras particulares (o en todas las interpretaciones sobre todos los dominios). Dependiendo de cuáles y cuántos conjuntos haya, habrá más o menos sentencias lógicamente válidas. Cuantos más conjuntos haya, tantas menos fórmulas serán satisfechas en todos ellos, es decir, tantas menos fórmulas (y en especial sentencias) serán lógicamente válidas.

Empecemos por una cuestión trivial. Normalmente no admitimos en nuestra semántica (o teoría de modelos) sistemas o estructuras con universo vacío. Debido a dicha convención, fórmulas como las siguientes son lógicamente válidas:

$$\exists x\,(Px \vee \neg Px) \qquad \exists x\; x = x \qquad \forall x\, Px \Rightarrow \exists x\, Px$$

Si adoptamos la convención contraria y admitimos estructuras particulares con universo vacío, entonces esas fórmulas dejan de ser lógicamente válidas.

Algunas fórmulas de primer orden (llamadas por Quine esquemas de infinitud) sólo son satisfechas por estructuras infinitas (es decir, con universo infinito), como las siguientes:

$$\forall x \exists y\, Rxy \wedge \forall x\, \neg Rxx \wedge \forall xyz\, (Rxy \wedge Ryz \Rightarrow Rxz)$$

$$\forall x \exists y \forall z\, (Rxy \wedge \neg Rxx \wedge (Ryz \Rightarrow Rxz))$$

[la misma, en forma normal prenexa]

$$\forall xy\, (f(x) = f(y) \Rightarrow x = y) \wedge \exists y \forall x\, f(x) \neq y$$

Las negaciones de estas fórmulas, como

$$\neg \forall x \exists y \forall z\ (Rxy \wedge \neg Rxx \wedge (Ryz \Rightarrow Rxz)),$$

son lógicamente verdaderas (o válidas) si todos los conjuntos son finitos, pero no son válidas si hay conjuntos infinitos.

Muchas fórmulas de primer orden que no son válidas si hay conjuntos infinitos (como normalmente suponemos que hay) pasan a ser lógicamente válidas si restringimos nuestro contexto conjuntista a los conjuntos finitos. La fórmula

$$\forall xyz\ (Rxy \wedge Ryz \Rightarrow Rxz) \wedge \forall x\ \neg Rxx \wedge \forall xy\ (Rxy \vee Ryx \vee x = y) \\ \Rightarrow \exists y \forall x\ (Rxy \vee x = y)$$

dice que si R es un orden lineal, entonces R tiene un máximo. Si hay conjuntos infinitos, entonces algunos órdenes lineales tienen un máximo y otros órdenes lineales carecen de máximo. En ese caso, la fórmula resulta verdadera (o satisfecha) en ciertas estructuras (los órdenes lineales con máximo) y falsa (o no satisfecha) en otras estructuras (los órdenes lineales sin máximo), por lo que no es lógicamente válida. Por el contrario, si no hay conjuntos infinitos, si todos los conjuntos son finitos, entonces cada orden lineal tiene un máximo, por lo que la fórmula es lógicamente válida.

La fórmula de primer orden

$$\forall xyz\ (Rxy \wedge Ryz \Rightarrow Rxz) \wedge \forall x\ \neg Rxx \wedge \forall xy\ (Rxy \vee Ryx \vee x = y) \\ \Rightarrow \exists xy\ (Rxy \wedge x \neq y) \wedge \neg \exists z\ (Rxz \wedge Rzy)$$

dice que si R es un orden lineal, entonces R no es denso. Desde luego, ningún orden lineal finito es denso. Por tanto, si sólo hay conjuntos finitos, esta fórmula es lógicamente verdadera. Sin embargo, si hay conjuntos infinitos, esta fórmula no es lógicamente válida, pues muchos órdenes lineales infinitos (como el de los racionales o el de los reales) son densos.

La fórmula de primer orden

$$\forall xyz\ (Rxy \wedge Ryz \Rightarrow Rxz) \wedge \forall x\ \neg Rxx \wedge \forall xy\ (Rxy \vee Ryx \vee x = y) \\ \Rightarrow [\forall xy\ (Sxy \Rightarrow Rxy) \wedge \exists xy\ Sxy \Rightarrow \exists z \forall xy\ (Sxy \vee Syx \Rightarrow Szx \vee z = x)]$$

dice que si R es un orden lineal y S es un suborden no vacío de R, entonces S tiene un mínimo. Su validez implicaría que cualquier orden lineal es un buen orden. En efecto, cualquier orden lineal finito es un

buen orden, y así la fórmula es lógicamente válida si todos los conjuntos son finitos. Pero muchos órdenes lineales infinitos (como el de los números enteros o el de los racionales o el de los reales) no son buenos órdenes. Por tanto, la fórmula no es lógicamente válida si hay conjuntos infinitos.

En contraste con la validez lógica de segundo orden, que no está bien definida, la validez de primer orden está bien determinada, tanto si aceptamos conjuntos infinitos como si nos limitamos a los conjuntos finitos. Sin embargo, en ambos casos obtenemos nociones distintas de validez lógica y conjuntos diferentes de sentencias lógicamente válidas. De todos modos, las sentencias válidas en la lógica estándar de primer orden, que admite conjuntos infinitos, siguen siendo válidas en la lógica finitaria de primer orden, que no los admite.

Si rechazamos los conjuntos infinitos y admitimos sólo conjuntos finitos, obtenemos muchas más fórmulas lógicamente válidas de primer orden. Puesto que hay menos conjuntos, y por tanto menos estructuras particulares, habrá más fórmulas satisfechas por todas las estructuras. Todas las fórmulas normalmente consideradas como válidas continuarán siendo válidas, pero otras muchas fórmulas nuevas pasarán a ser válidas, como las que acabamos de considerar.

Hemos visto que la lógica de segundo orden depende masivamente de la teoría de conjuntos, pero también la lógica de primer orden tiene una cierta dependencia. En la polémica sobre el logicismo mantenida en las primeras décadas del siglo XX, el *status* del axioma de infinitud desempeñó un papel crucial. La polémica concluyó con el acuerdo de que el axioma de infinitud pertenece a la teoría de conjuntos y no tiene nada que ver con la lógica. Sin embargo, y como acabamos de ver, tiene mucho que ver con la lógica, incluso ya con la lógica de primer orden. Según que lo aceptemos o no en nuestra metateoría como contexto de nuestras definiciones, obtenemos lógicas de primer orden completamente (extensionalmente) diferentes.

La noción estándar de lógica de primer orden acepta la existencia de conjuntos infinitos en el contexto conjuntista de la metateoría. Esta lógica (la usual) es semánticamente completa, como Gödel probó en 1930. Esto significa que el conjunto de las sentencias válidas de la lógica de primer orden es recursivamente enumerable; en otras palabras, que puede ser generado por la aplicación sucesiva de las reglas de un cálculo deductivo. Esto equivale también a decir que el cálculo nos permite deducir todas las consecuencias de un conjunto dado de premisas. Todas las pruebas (debidas a Gödel, Henkin y otros) de la completud semántica de la lógica de primer orden se basan en la construcción de ciertos conjuntos infinitos (como cierto conjunto in-

finito de términos, en el caso de Henkin) y por tanto se vienen abajo si rechazamos los conjuntos infinitos. Además, si rechazamos los conjuntos infinitos, reducimos tan drásticamente la cantidad de estructuras disponibles que el número de fórmulas lógicamente válidas aumenta dramáticamente. De hecho, el conjunto de las fórmulas válidas se torna tanto mayor y más complejo, que incluso deja de ser recursivamente enumerable, como probó Trakhtenbrot en 1950. En resumen, la lógica de primer orden sólo es semánticamente completa en la medida en que admitamos conjuntos infinitos en nuestro contexto metateórico.

REFERENCIAS BIBLIOGRÁFICAS

Cohen, P. (1963-64), «The Independence of the Continuum Hypothesis, I and II»: *Proceedings of the National Academy of Sciences USA*, 50, 1143-1148 y 51, 105-110.

Gödel, K. (1930), «Die Vollständigkeit der Axiome des logischen Funktionenkalküls»: *Monatshefte für Mathematik und Physik*, 37, 349-360.

Gödel, K. (1931), «Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme»: *Monatshefte für Mathematik und Physik*, 38, 173-198.

Gödel, K. (1940), *The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis with the Axioms of Set Theory*, Princeton University Press.

Gödel, K. (1947), «What is Cantor's Continuum Problem?»: *The American Mathematical Monthly*, 54, 515-525.

Henkin, L. (1950), «Completeness in the Theory of Types»: *Journal of symbolic Logic*, 15, 81-91.

Jané, I. (1993), «A Critical Appraisal of Second-Order Logic»: *History and Philosophy of Logic*, 14, 67-86.

Jané, I. (1995), «Lógica de orden superior», en *Enciclopedia Iberoamericana de Filosofía*, vol. 7, *Lógica*, Trotta, Madrid, 105-128.

Kanamori, A. (1996), «The Mathematical Development of Set Theory from Cantor to Cohen»: *The Bulletin of Symbolic Logic*, vol. 2, 1-71.

Moore, G. (1980), «Beyond First-Order Logic: the Historical Interplay between Logic and Set Theory»: *History and Philosophy of Logic*, vol. 1, 95-137.

Mosterín, J. (1987), «La frontera entre lógica y teoría de conjuntos», en *Symposium Quine: Actas del Symposium Internacional sobre el Pensamiento Filosófico de W. V. Quine*, Granada, 209-223.

Mosterín, J. (2000), *Los lógicos*, Espasa, Madrid.

Mosterín, J. (2002), «How Set Theory Impinges on Logic», en E. Agazzi y P. Weingartner (ed.), *Alternative Logics: Do Sciences Need Them?*, Springer-Verlag, New York.

Mosterín, J. y Torretti, R. (2002), *Diccionario de Lógica y Filosofía de la Ciencia*, Alianza, Madrid.
Shapiro, S. (1990), «Second-Order Logic, Foundations and Rules»: *The Journal of Philosophy*, vol. 87, 234-261.
Trakhtenbrot, B. A. (1950), «The impossibility of an algorithm for the decision problem in finite classes [en ruso]»: *Dokl. Akad. Nauk SSSR*, 70, 569-572.

¿DE QUÉ TRATA LA TEORÍA DE CONJUNTOS?

Ignacio Jané

Este artículo es una reflexión filosófica sobre la teoría de conjuntos. Se divide naturalmente en dos partes, la primera de las cuales, que comprende las seis primeras secciones, es una presentación de la teoría de conjuntos actual como la articulación matemática de la llamada «concepción iterativa». Esta concepción es sólo esquemática, pues si bien describe el universo de los conjuntos como el resultado de la iteración absolutamente ilimitada de la operación que asigna a cada dominio de objetos la totalidad de los conjuntos de estos objetos, deja sin explicar el alcance de la ilimitación absoluta de la iteración y, más fundamentalmente, el concepto de conjunto arbitrario de objetos dados que está en su base. La segunda parte del articulo, constituida por las tres últimas secciones, trata de la interpretación filosófica de la teoría de conjuntos. Si bien esta teoría suele verse como una encarnación del realismo en matemáticas, hay que distinguir (y la distinción es crucial) entre el realismo meramente metodológico, que sí debemos atribuir a la teoría de conjuntos (o a su uso), y el realismo propiamente filosófico sobre el universo conjuntista, que podemos tratar de rehuir sin por ello revisar la teoría. En la última sección esbozamos y motivamos un modo no realista de concebir la base en que descansa la concepción iterativa, esto es, el paso de un dominio de objetos al dominio de todos los conjuntos de estos objetos. Este paso, si el dominio es infinito, es de una enorme complejidad.

1. LA TEORÍA DE CONJUNTOS COMO TEORÍA DEL INFINITO ACTUAL

Si, como escribió Hermann Weyl (1925), la matemática es la ciencia del infinito, el objeto de la teoría de conjuntos es el infinito actual. Hasta el advenimiento de la teoría de conjuntos en la segunda mitad del siglo XIX, las distintas disciplinas matemáticas se ocupaban de infinitos objetos y de sus relaciones mutuas, pero no de objetos infinitos: los objetos de estudio de la matemática (números, figuras, magnitudes) eran, ellos mismos, finitos. Por otra parte, la infinitud de su multiplicidad era tratada como potencial. Si, hablando en sentido potencial, decimos que hay infinitos objetos de cierto tipo, lo que estamos diciendo es que ninguna colección finita contiene todos los objetos en cuestión, lo cual no equivale a afirmar que hay una colección infinita que los contiene todos[1]. De modo análogo, la afirmación de que un número irracional, como π o como $\sqrt{2}$, es el límite de una sucesión infinita de aproximaciones racionales, no debe entenderse como la afirmación de la existencia de una sucesión completa de términos racionales cuyas diferencias con π o con $\sqrt{2}$ son cada vez menores, sino más bien como la existencia, o la posibilidad de obtención, para cualquier entero positivo n, de un número racional cuya diferencia con π o con $\sqrt{2}$ es menor que $1/n$; los términos de la sucesión no existen, por así decir, todos simultáneamente, pero es posible generar cualquier número prefijado de ellos. En otras palabras, es posible obtener aproximaciones racionales tan exactas a π o a $\sqrt{2}$ como queramos, pero no se pretende que exista ninguna lista completa de las aproximaciones.

Esta concepción de la aproximación como un proceso no presenta dificultades importantes cuando el número irracional en cuya aproximación nos interesamos está dado de antemano (como π y $\sqrt{2}$, que aparecen en consideraciones geométricas), pero muestra sus deficiencias cuando tratamos de precisar con su ayuda qué es un número real arbitrario. Aquí invertimos la relación; en vez de partir del número real y obtener después la sucesión de aproximaciones, partimos de una sucesión apropiada de números racionales y determinamos el número real a partir de ella. Pero entonces, sea lo que sea el número que queremos determinar, está claro que para hacerlo necesitamos apelar a la totalidad de los términos de la sucesión, puesto que una misma

1. Compárese con la proposición 20 del libro IX de los *Elementos* de Euclides: «Hay más números primos que cualquier cantidad propuesta de números primos». Según Aristóteles, «lo infinito es aquello fuera de lo cual siempre hay algo» (*Física*, 206b 33).

sucesión finita de números racionales, por larga que sea, es una aproximación tan precisa como queramos de una infinitud de números reales.

Un modo especialmente perspicuo de precisar el concepto de número real lo hallamos en Dedekind (1872): partiendo de la observación de que los números racionales menores que un número real dado constituyen un segmento inicial propio sin elemento máximo, definimos un número real precisamente como un segmento inicial propio de números racionales. Los números reales son, pues, conjuntos infinitos de números racionales. De este modo, a través del análisis entra en la matemática el infinito actual, el infinito que es característico de los conjuntos, ya que no tiene sentido decir que un conjunto posee infinitos elementos en potencia.

Cantor hizo de los conjuntos infinitos objeto propio de estudio matemático, como sugiere el título común de los seis artículos que constituyen su obra fundamental: «Sobre conjuntos infinitos de puntos lineales» (*Über unendliche lineare Punktmannigfaltigkeiten*, en Cantor [1932], 139-246). Cantor trata los conjuntos infinitos como hasta entonces sólo se trataban los conjuntos finitos: opera con ellos, transforma unos en otros mediante reglas precisas, los compara respecto a su cardinalidad y muestra cómo asignar un número cardinal a cada conjunto. Su primer resultado fue mostrar que el conjunto de los números naturales y el de los números reales tienen cardinalidad distinta[2].

2. LA TEORÍA DE CONJUNTOS COMO FUNDAMENTO DE LA MATEMÁTICA

Por una parte, la teoría de conjuntos es una teoría matemática entre otras teorías matemáticas, con un objeto de estudio propio y con métodos propios, con relaciones más o menos profundas con otras teorías matemáticas. Por otra parte, sin embargo, la teoría de conjuntos es una teoría matemática peculiar, en cuanto la mayor parte del resto de la matemática, en particular todas las teorías matemáticas tradicionales, son interpretables en ella. Que una teoría T sea interpretable en la teoría de conjuntos significa que es posible tratar los objetos

2. Dos conjuntos tienen la misma cardinalidad si son biyectables, es decir, si hay una correspondencia biunívoca entre ellos; si dos conjuntos no tienen la misma cardinalidad pero uno es biyectable con un subconjunto de otro, la cardinalidad del primero es menor que la del segundo. Estas definiciones se deben a Cantor.

de que *T* se ocupa como conjuntos, y los conceptos, las operaciones y las relaciones que le son propias como conceptos de conjuntos, operaciones con conjuntos y relaciones entre conjuntos, y ello de modo tal que a cada una de las proposiciones expresables en el lenguaje de *T* se le asocia de manera sistemática una proposición conjuntista y que las proposiciones conjuntistas asociadas a los teoremas de *T* son teoremas de la teoría de conjuntos. Brevemente, interpretar una teoría matemática en la teoría de conjuntos equivale a reformularla como un fragmento de la teoría de conjuntos.

La interpretación de la matemática en la teoría de conjuntos tiene por lo menos dos consecuencias importantes, una de posible uso filosófico y otra con efectos en la investigación matemática no conjuntista. La primera es que es posible desplazar la discusión de problemas ontológicos de las diversas teorías matemáticas a la teoría de conjuntos. El caso de los números reales es un ejemplo paradigmático: los problemas filosóficos acerca de la existencia de los números reales pueden retrotraerse a la existencia de suficientes conjuntos de números racionales. La segunda consecuencia es que la interpretabilidad de una teoría matemática *T* en la teoría de conjuntos permite mostrar que ciertos problemas abiertos en *T* son insolubles, pues si la traducción conjuntista de una proposición expresable en el lenguaje de *T* es independiente de los principios básicos de la teoría de conjuntos, es decir, no es demostrable ni refutable a partir de ellos, entonces la proposición en cuestión tampoco es demostrable ni refutable en *T*. Esta consecuencia es importante a causa del gran desarrollo de los métodos de demostración de independencia en teoría de conjuntos.

Es habitual referirse al hecho de que buena parte de la matemática es interpretable, o reducible, como también se dice, a la teoría de conjuntos, hablando de esta teoría como de un *fundamento* de la matemática. Podemos hacerlo si no damos a este término un significado más fuerte. Hablar de fundamento en esta situación no supone que la teoría de conjuntos confiera a las teorías interpretables en ella ningún carácter especial de solidez, de seguridad frente a la contradicción, o de evidencia (de hecho, en ciertos casos, como la teoría de números es obvio lo contrario). Lo único que significa es que puede tomarse como base, que es posible en principio reelaborar el resto de la matemática en términos de conjuntos y a partir de principios sobre existencia, estructura y relaciones mutuas entre ellos.

Un tercer aspecto de la teoría de conjuntos es su uso generalizado como herramienta en la matemática contemporánea. Aunque los objetos de las distintas teorías matemáticas no se conciban como conjuntos, se consideran conjuntos de estos objetos, se opera con estos

conjuntos y se aplican técnicas conjuntistas para extraer información acerca de ellos. La ubicuidad de las construcciones y métodos conjuntistas en teorías matemáticas muy dispares confiere a la matemática una unidad de la que carecía hasta finales del siglo XIX y que adquirió con el uso progresivo de la teoría de conjuntos.

3. EXTENSIONES DE CONCEPTOS

La mera interpretabilidad de una teoría matemática en la teoría de conjuntos no garantiza su certeza, ni su aprioridad, ni cualquier otra propiedad que pudiéramos considerar deseable desde una perspectiva epistemológica o, más generalmente, filosófica. Sólo podría garantizarlo si la teoría de conjuntos tuviera estos atributos. Si alguna vez se ha pensado que los tenía es porque se consideraba a los conjuntos como entidades lógicas, idénticas, o, por lo menos, análogas a las extensiones de los conceptos fregeanos. Si lo fueran, cabría dar a la palabra «fundamento» el significado más pleno que sugiere.

Las investigaciones lógico-matemáticas de Frege presuponen dos tipos de entidades, los objetos y las funciones. Entre las funciones se cuentan los conceptos, que asignan un valor de verdad a cada objeto: lo *verdadero*, si el objeto cae bajo el concepto, lo *falso* si no cae bajo él (los valores de verdad son objetos). Si bien los conceptos no son objetos, a cada concepto le corresponde un objeto, su *extensión*, que podemos concebir como la totalidad de los objetos que caen bajo el concepto, de modo que la relación fundamental entre conceptos y extensiones la expresa el famoso axioma V de *Las leyes fundamentales de la aritmética*, según el cual las extensiones de dos conceptos son iguales si y sólo si bajo ambos conceptos caen exactamente los mismos objetos[3].

De acuerdo con Frege, la relación de concepto a extensión es una relación lógica y la extensión de un concepto lógico es un objeto lógico. Así, una teoría general de las extensiones de conceptos será una teoría lógica, y si a ella logramos reducir una teoría matemática, podremos mantener que hemos mostrado su carácter lógico, fundamentándola en un sentido fuerte.

Frege logró interpretar la aritmética de los números naturales en una teoría tal, que en retrospectiva podemos describir como un sistema de lógica de segundo orden ampliado con el axioma V. Pero esta

3. De hecho, el axioma V es más general, ya que habla de funciones cualesquiera, no sólo de conceptos.

teoría es contradictoria, como puso de manifiesto Bertrand Russell. Para verlo, consideremos el concepto *F* bajo el cual, por definición, cae un objeto *x* si y sólo si *x* es la extensión de algún concepto bajo el cual *x* no cae[4]. Sea a la extensión de *F* y preguntémonos si *a* cae o no cae bajo *F*. Si *a* cae bajo *F*, hay un concepto *G* tal que *a* es la extensión de *G* y *a* no cae bajo *G*; pero entonces, puesto que la extensión de *G* es la de *F* y *a* no cae bajo *G*, el axioma V nos permite concluir que *a* no cae bajo *F*. Si, por otra parte, *a* no cae bajo *F*, hay ciertamente un concepto (a saber, *F*), cuya extensión es *a* y bajo el cual *a* no cae, de modo que, por definición de *F*, *a* cae bajo *F*. En consecuencia,

a cae bajo *F* si y sólo si *a* no cae bajo *F*,

lo cual es una contradicción.

4. SOBRE EL PRINCIPIO DE COMPRENSIÓN

La contradicción de Russell se hace más perspicua en el contexto de lo que se ha llamado la *teoría ingenua de conjuntos*, claramente emparentada con la teoría de Frege y basada en dos principios: el de extensionalidad, que dice que conjuntos con los mismos elementos son iguales, y el de comprensión, según el cual toda propiedad determina un conjunto, en el sentido de que existe un conjunto cuyos elementos son exactamente los objetos que tienen la propiedad en cuestión. La contradicción se obtiene considerando el conjunto *a* determinado por la propiedad de ser un conjunto que no es miembro de sí mismo, o sea, que *a* es el conjunto cuyos elementos son los conjuntos que no son elementos de sí mismos. Concluimos que *a* es un elemento de *a* si y sólo si *a* no es un elemento de *a*.

Se ha mantenido a menudo que la teoría ingenua de conjuntos es la concepción natural y que cualquier restricción del principio ilimitado de comprensión, el responsable de la contradicción de Russell, es más o menos arbitraria[5]. Sin embargo, el supuesto carácter evidente

4. En símbolos, $Fx \leftrightarrow \exists G\ (x = \mathrm{ext}(G) \wedge \neg Gx)$. Esta definición de *F* es admisible en el sistema de Frege.

5. Así, leemos en Quine, 1976: «La paradoja de Russell es una antinomia genuina, puesto que el principio de existencia de clases que nos fuerza a abandonar es fundamental» (12). Más adelante, Quine afirma que, para evitar la contradicción, «nos vemos obligados a buscar combinaciones óptimas de supuestos de existencia y, en consecuencia, hay una gran variedad de propuestas para los fundamentos de la teoría general de conjuntos. Cada una de las propuestas es innatural, puesto que el esquema natural es el irrestricto que las antinomias desacreditan» (16).

del principio ilimitado de comprensión, el responsable de la paradoja, es ilusorio; la versión intuitivamente aceptable es más restringida.

Supongamos que partimos de ciertos objetos y de una condición C con respecto a la cual está determinado, para cada uno de estos objetos, si cumple la condición o no. El principio ilimitado de comprensión afirma que hay una clase, o un conjunto, cuyos elementos son exactamente los objetos que cumplen la condición[6].

Nos preguntamos cómo podría no existir esta clase, si hemos sido capaces de definirla. Debe existir, decimos, puesto que para garantizar la existencia de una clase basta que esté determinado cuáles son sus elementos. Y de la supuesta clase que se obtiene por comprensión está bien determinado cuáles son: son los objetos que cumplen la condición C.

Pero aquí nos hemos precipitado. Hemos pasado de la suposición de que para cada objeto está determinado si cumple la condición *C* a la conclusión de que, por tanto, está determinado cuáles son todos los objetos que la cumplen. Para obtener esta conclusión necesitamos una premisa adicional: que esté determinado qué objetos hay, es decir, cuáles son todos los objetos.

Ésta es una premisa implícita y esencial en la justificación intuitiva del principio de comprensión. El principio de comprensión intuitivamente motivado sólo se aplica a dominios bien determinados de objetos: de antemano está determinado qué objetos hay; por tanto, si está determinado de cada uno de ellos si cumple la condición C, ha de estar determinado cuáles son los objetos que cumplen la condición. En este caso, pues, la clase determinada por C existe, aunque puede no ser uno de los objetos de que partimos (no lo será si *C* es la condición de Russell).

Las clases tienen límites definidos, y si el ámbito del que partimos no los tiene, en cualquier separación de las entidades consideradas en dos partes, por preciso que sea el criterio seguido, por lo menos una no será una clase.

No sólo el principio ilimitado de comprensión no es evidente, sino que su rechazo es parte de una tradición distinguida del pensamiento occidental, como lo atestigua la aceptación del infinito potencial junto a la negación del infinito actual. En efecto, quien, admitien-

6. Dos acotaciones terminológicas: 1) Una condición describe una propiedad, de modo que podemos decir que un objeto tiene cierta propiedad o que cumple cierta condición. No establecemos ninguna diferencia entre ambas formas de expresión; usamos una u otra por razones estilísticas o de tradición. 2) Tomamos «clase» y «conjuntos» como sinónimos, si bien aquí hablamos de clases, ya que es el término usado por Russell.

do que el concepto de número natural es preciso, mantiene que los números naturales son potencialmente infinitos pero no lo son en acto, está negando (aunque no lo exprese de este modo) que la condición de ser un número natural determina un conjunto. No es que niegue que la totalidad de los números naturales sea concebible como un objeto, sino que niega la existencia de esta totalidad; niega, dicho de un modo sugerente, que los números naturales estén todos disponibles para constituir un conjunto o una clase.

Así, lo que la contradicción de Russell muestra es que no está determinado qué conjuntos hay, o, como diremos en adelante, que los conjuntos no constituyen un dominio. En general, cuando digamos que ciertos objetos forman un *dominio*, significaremos con ello que está determinado cuáles son exactamente estos objetos; no sólo si un objeto cualquiera es uno de los objetos en cuestión, sino cuáles son exactamente todos los objetos en cuestión.

5. LA CONCEPCIÓN ITERATIVA

Para Russell cuando descubrió su paradoja, como más tarde para Quine, el principio de comprensión era el principio natural sobre existencia de clases o conjuntos. Pero no lo era para Cantor, el fundador de la teoría de conjuntos. Cantor no concebía un conjunto como una parte del universo segregada por una condición o una propiedad, sino como una parte de un dominio, o, en sus propias palabras, de una esfera conceptual (*Begriffssphäre*)[7]. Ya antes del descubrimiento de cualquier contradicción Cantor negó que toda propiedad determinara un conjunto. La contradicción de Russell (a la que, dicho sea de paso, no llegó como resultado del análisis de la obra de Frege, sino como un intento de aplicar el método diagonal de Cantor a la clase universal[8])

7. Como hemos dicho en la primera sección, la serie de artículos en que Cantor desarrolló su teoría trata de conjuntos infinitos de puntos lineales, o sea, de conjuntos infinitos de puntos del dominio constituido por la recta real.

8. Con cierta imprecisión histórica que no afecta al contenido, Cantor introdujo su método diagonal para mostrar que la cardinalidad del conjunto potencia de un conjunto es mayor que la del conjunto mismo, en otras palabras (ver nota 2), que, dado un conjunto a, no hay ninguna función con dominio a cuyos valores sean todos los subconjuntos de a. La demostración es constructiva: si f es una función que asigna un subconjunto de a a cada elemento de a, el conjunto $b = \{x \in a : x \notin f(x)\}$, explícitamente definido en términos de a y de f, no es un valor de f (pues para todo $x \in a$, $x \in b$ si y sólo si $x \notin f(x)$). La clase contradictoria de Russell se obtiene tomando como a la clase universal y como f la función definida por $f(x) = \{x\}$, si x no es una clase y $f(x) = x$, si x es una clase. Vemos que, para todo $x \in a$, $x \notin f(x)$ si y sólo si x es una clase, y

no afectó a Cantor, que la vio como el resultado de una aplicación ilegítima de sus métodos (véase la carta de Cantor a Jourdain del 9 de julio de 1904, en Grattan-Guinness, 1971). De hecho, quien primero descubrió el argumento de la paradoja de Russell fue Zermelo, el autor de la primera axiomatización de la teoría de conjuntos de Cantor, que no vio en ella una contradicción, sino una prueba formal de que no hay ningún conjunto universal (véase Rang y Tomas, 1981).

El concepto de conjunto de la teoría actual es el cantoriano: un conjunto es una colección de objetos de un dominio (más adelante precisaremos y matizaremos esta afirmación). Pero esto no es todo, naturalmente, ya que debemos aclarar cómo se obtienen los dominios. La idea del universo conjuntista que descubrimos en la teoría actual es la de una jerarquía de dominios cada vez más incluyentes y de los conjuntos en ellos incluidos. Esta idea no se halla en Cantor, pero sus ingredientes son cantorianos.

La suposición fundamental es que a cada dominio D le corresponde un nuevo dominio, $\mathcal{P}(D)$, el dominio potencia de D, que consta de todos los conjuntos de elementos de D. Obtenemos los dominios iterando la operación $D \rightarrow \mathcal{P}(D)$ a partir de un dominio básico, uno cuya condición de dominio es evidente: el dominio vacío. Así, partiendo de $D_0 = \varnothing$, obtenemos los dominios $D_1 = \{\varnothing\}$, $D_2 = \{\varnothing, \{\varnothing\}\}$, ..., de modo que, para cada n, $D_{n+1} = \mathcal{P}(D_n)$. Pero la iteración va más allá. Tras la sucesión de los D_n obtenemos un nuevo dominio, cuyo contenido no es más que la unión de los dominios anteriores, y podemos formar su conjunto potencia.

Ahora bien, para poder continuar la iteración más allá de lo finito, debemos disponer de índices que extiendan la sucesión de los números naturales. Estos índices son los ordinales transfinitos que Cantor introdujo en 1883 y que cumplen un papel central en la teoría de conjuntos. Cantor define los números ordinales mediante un proceso de generación a partir del número cero (el menor ordinal), con ayuda de dos principios: dado un ordinal a, el primer principio permite generar su sucesor inmediato $\alpha + 1$, mientras que el segundo principio se aplica a cualquier sucesión de ordinales sin elemento máximo, produciendo su límite, es decir, el menor ordinal mayor que todos los términos de la sucesión. Los primeros ordinales, los ordinales finitos, son los números naturales: 0, 1, 2, ..., que se obtienen aplicando sólo el primer principio. Éstos forman una sucesión sin elemento máximo, a la que el segundo principio se aplica, obteniendo el menor ordinal

$x \notin x$, por lo que el conjunto b es la clase de Russell: $b = \{x : x \text{ es una clase y } x \notin x\}$ (véase Cantor, 1932, 278-271, y Russell, 1903, §§ 346-349).

infinito o transfinito, el número ω, el cual es, a su vez, el primer término de una nueva sucesión de ordinales obtenidos mediante aplicaciones sucesivas del primer principio: ω, $\omega + 1$, $\omega + 2$, ... Esta sucesión no tiene elemento máximo, por lo que a ella es aplicable el segundo principio, que da lugar al ordinal $\omega + \omega$, o $\omega \cdot 2$, al cual se aplica también el primer principio, etc. Con ayuda de ambos principios se obtienen progresivamente $\omega \cdot 3$ (que es el límite de la sucesión $\omega \cdot 2 + 1$, $\omega \cdot 2 + 2$, ...), $\omega \cdot 4$, ..., $\omega \cdot n$, ..., su límite $\omega \cdot \omega$, o ω^2. Se obtienen también ω^3, ω^4, ..., su límite ω^ω, etcétera.

De la descripción del modo de operar de estos dos principios se sigue que la generación de ordinales es un proceso totalmente incompletable: no puede haber ninguna colección no vacía de ordinales que esté cerrada con respecto a los dos principios de generación. La razón es que toda colección de ordinales forma naturalmente una sucesión y siempre hay un ordinal mayor que cualquier sucesión de ordinales: si la sucesión dada tiene elemento máximo, el primer principio nos proporciona un ordinal tal, si no tiene elemento máximo, nos lo proporciona el segundo. En palabras de Cantor: «el segundo principio [...] se presenta como un medio que, junto con el primer principio de generación, ofrece la capacidad de superar toda barrera en la formación conceptual de los números [ordinales]» (Cantor, 1932, 196-197). Cantor veía la generación de los ordinales como absolutamente infinita:

> No me cabe duda alguna de que por este camino nunca hallaremos un límite que no sea posible franquear, pero tampoco obtendremos una concepción siquiera aproximada del absoluto. El absoluto sólo puede ser admitido, pero nunca conocido, ni siquiera aproximadamente [...]. La sucesión absolutamente infinita de los números me parece, pues, en cierto sentido, un símbolo adecuado del absoluto (Cantor, 1932, 205 n. 2).

Ni para Cantor ni para la teoría de conjuntos actual los ordinales constituyen un conjunto. La propiedad de ser un ordinal fue el primer contraejemplo que Cantor dio al principio de comprensión, ocho años antes de que Russell descubriera su paradoja.

Con la ayuda de los ordinales, podemos dar una descripción esquemática del universo de la teoría de conjuntos. Los conjuntos se distribuyen en estratos sucesivos, indicados por los ordinales. El estrato, o universo parcial, correspondiente al ordinal a es el dominio V_α, que se obtiene a partir de sus predecesores, siguiendo el orden de generación de los ordinales. Puesto que el primer estrato, V_0, es vacío, los conjuntos que obtenemos en los distintos estratos son con-

juntos puros[9]. Si α es un ordinal distinto cero, definimos V_α de uno u otro modo según como *a* haya sido generado: si *a* se ha obtenido por el primer principio, de modo que es el sucesor inmediato de un cierto ordinal β ($\alpha = \beta+1$), V_α es el dominio potencia de V_β; si, por el contrario, α ha sido obtenido por el segundo principio, si es el límite de una sucesión de ordinales sin elemento máximo, el estrato V_α es la unión de todos los estratos precedentes, de modo que en V_α no aparece nada nuevo, simplemente se acumulan todos los conjuntos ya presentes en los estratos anteriores. Observemos que también en este segundo caso V_α es un dominio; en efecto, puesto que está determinado cuáles son exactamente todos los ordinales menores que α y qué conjuntos contiene cada uno de los dominios V_γ, para $\gamma < \alpha$, está también determinado qué conjuntos contiene exactamente V_α.

De esta descripción esquemática se sigue que los estratos son acumulativos, es decir, si un conjunto aparece en un estrato, aparece también en todos los estratos ulteriores, si bien hay siempre un primer estrato en el que un conjunto dado aparece. Además, los elementos de un conjunto aparecen en estratos anteriores al primero en que el conjunto aparece. En razón de estas dos características, nos referimos a la sucesión de estratos como la *jerarquía acumulativa* y a la idea subyacente como la *concepción iterativa* de los conjuntos[10].

La concepción iterativa no es fruto de un intento de evitar las paradojas, sino que es una articulación elaborada de la idea natural de que para formar un conjunto hemos de disponer de sus posibles elementos, en otras palabras, de que los elementos de un conjunto son previos al conjunto. Los objetos previos son los objetos de un dominio; los conjuntos son los elementos del correspondiente dominio potencia. El paso de un dominio a su dominio potencia es suficiente para muchas aplicaciones puntuales (es, en esencia, la base de la lógica de segundo orden), pero en teoría de conjuntos los conjuntos son los objetos de estudio, lo cual da cuenta de la necesidad de la iteración. Que la iteración sea ilimitada (a lo largo de los ordinales) es la expresión del deseo de que toda posible colección esté adecuadamente representada por un conjunto. Para que esto sea así, el universo conjuntista debe ser abierto, no puede ser un dominio y, por tanto, no puede ser un conjunto. En la concepción iterativa el universo

9. A partir de un conjunto a obtenemos el conjunto formado por elementos de *a*, los elementos de los elementos de *a*, los elementos de estos elementos, etc.; este conjunto es la *clausura transitiva de a*. Un conjunto es *puro* si su clausura transitiva contiene sólo conjuntos; en otro caso es un conjunto *impuro*.

10. Dos presentaciones de la concepción iterativa no del todo coincidentes con la que ofrecemos aquí se hallan en Boolos, 1971, y en Potter, 1993.

es abierto porque la generación de los ordinales lo es, y la generación de los ordinales es abierta porque, como hemos visto, las reglas de generación no admiten clausura.

Es apropiado concluir la presentación de la concepción iterativa con estas palabras de Gödel:

> [La paradojas] son un problema muy serio, pero no para la teoría de conjuntos de Cantor [...] [El concepto cantoriano de conjunto], según el cual un conjunto es cualquier cosa obtenible a partir de los enteros (o de otros objetos bien definidos) por aplicación iterada de la operación «conjunto de», y no algo obtenido dividiendo la totalidad de todas las cosas que existen en dos categorías, nunca ha dado lugar a ninguna antinomia; es decir, el trabajo perfectamente «ingenuo» y acrítico con este concepto de conjunto ha resultado ser completamente autoconsistente (Gödel, 1947, 180).

6. LOS AXIOMAS DE ZF

Una lectura, incluso atenta y cuidadosa, de la lista de los axiomas de la teoría de conjuntos no sugiere ninguna conexión con la concepción iterativa. Los axiomas no hablan de ordinales ni de estratos, sino que, con sólo dos excepciones, son principios de existencia de conjuntos. A primera vista, los axiomas no son más que una serie de principios útiles, seleccionados sin otro fin que el de garantizar las construcciones y los medios de demostración que el matemático necesita. Sin embargo, esta apariencia es engañosa, ya que a partir de estos axiomas es posible definir el concepto de ordinal y demostrar que los ordinales definidos se comportan como los ordinales cantorianos, y es posible definir la sucesión de estratos y demostrar que todo conjunto pertenece a algún estrato; en otras palabras, los axiomas implican que los conjuntos se distribuyen según la jerarquía acumulativa[11].

Veremos ahora que, inversamente, los axiomas pueden justificarse a partir de la concepción iterativa, por lo que la teoría que los axiomas habituales determinan, la teoría de Zermelo-Fraenkel (ZF), es una expresión en lenguaje matemático preciso de la concepción ite-

11. Véase, por ejemplo, Fraenkel y Bar-Hillel, Levy, 1973, 86-95. Este libro ya clásico contiene una autorizada discusión de los aspectos fundacionales y filosóficos de la teoría de conjuntos. Para una presentación cuidadosa de la teoría, puede verse Levy, 1979 y, a un nivel más elemental, Moschovachis, 1994 y Vaught, 1995. Desde una perspectiva filosófica, cabe destacar Lavine, 1994. Ferreirós, 1999 es una excelente presentación histórica de la teoría de conjuntos con énfasis en los aspectos conceptuales.

rativa[12]. Éste es un buen argumento en favor de la naturalidad de la teoría de ZF y, por tanto, en contra, de que esta teoría no es más que una respuesta a las paradojas. Es preciso decir, sin embargo, que la teoría de ZF no caracteriza por completo a los conjuntos, en cuanto que hay muchas proposiciones importantes que es incapaz de decidir. Pero tampoco la concepción iterativa que hemos descrito caracteriza propiamente a la jerarquía acumulativa, ya que no es más que un esquema, un andamiaje que sólo adquirirá cuerpo en la medida en que demos contenido a los dos parámetros de que depende: el paso de un dominio a su dominio potencia y el alcance de la ilimitación de la generación de ordinales. Si reflexionamos sobre estos dos puntos, descubriremos la dificultad de articular con precisión qué es un conjunto arbitrario de objetos de un dominio, y qué significa la infinitud absoluta de los ordinales.

No es éste el lugar de dar una formulación precisa de los distintos axiomas y de justificarlos con todo detalle. Nos limitaremos a describir su contenido y a esbozar su justificación a partir de la concepción iterativa.

El axioma más básico es el de *extensionalidad*, que dice que todo conjunto está determinado por sus elementos, es decir, que dos conjuntos distintos difieren en algún elemento. Su justificación, previa a la concepción iterativa, descansa en que los conjuntos se comportan como colecciones. Otro axioma es el de *fundación* o de *regularidad*, según el cual, si hay algún conjunto que posee cierta propiedad Φ, hay un conjunto minimal que la posee, es decir, hay un conjunto a que posee la propiedad Φ pero ningún elemento de a la posee. Su justificación reside en que de los estratos que contienen conjuntos con la propiedad Φ uno es el primero, por lo que ningún elemento de un conjunto de este primer estrato posee la propiedad Φ.

Los restantes axiomas son de existencia. Su justificación en la concepción iterativa consiste en mostrar que el conjunto cuya existencia el axioma afirma pertenece a algún estrato[13]. El axioma del *conjunto vacío* afirma la existencia de $\varnothing$; el axioma del *par* la del conjunto $\{a, b\}$, para cualesquiera conjuntos a y b; el axioma de la *unión* afirma que si a es un conjunto, hay un conjunto, $\bigcup a$, al que pertene-

12. Compárese la justificación que daremos con las que se hallan en Scott, 1974 y en Shoenfield, 1977.

13. Para verificar los detalles de la justificación que se esboza, es conveniente observar que de la descripción de la jerarquía acumulativa se sigue que un conjunto a pertenece al estrato V_α si y sólo si hay un estrato anterior V_β al cual todos los elementos de a pertenecen; en símbolos: $a \in V_\beta$ si y sólo si hay $\beta < \alpha$ tal que $\alpha \subseteq V_\beta$.

cen exactamente los elementos de los elementos de a; el axioma de *separación* afirma que si a es un conjunto y Φ es una propiedad, existe el conjunto $\{x \in a : \Phi(x)\}$ de los elementos de a que tienen la propiedad Φ; el axioma del *conjunto potencia*, que afirma si a es un conjunto, hay un conjunto $\mathcal{P}(a)$ cuyos elementos son todos los subconjuntos de a (es decir, todos los conjuntos de elementos de a). La justificación de estos axiomas en la concepción iterativa es simple: $\varnothing$ pertenece a V_1, y, si a está en V_α, b está en V_β, y $\alpha < \beta$, entonces $\{a, b\}$ está en $V_{\beta+1}$, $\bigcup a$ pertenece a V_α, al igual que $\{x \in a : \Phi(x)\}$, mientras que $\mathcal{P}(a)$ pertenece a $V_{\alpha+1}$.

Nos quedan dos axiomas de existencia más. El axioma de *infinitud* afirma la existencia de un conjunto infinito[14]. El conjunto V_ω lo es. El axioma de *substitución* dice que si asignamos a cada elemento x de un conjunto a un conjunto a_x, entonces hay un conjunto b al cual todos los a_x pertenecen (de manera sugestiva, aunque engañosa, que da razón del nombre: si «substituimos» los elementos de a por conjuntos cualesquiera, el resultado es también un conjunto). En el contexto de la jerarquía acumulativa, este axioma puede verse como un axioma fuerte de infinitud, en términos de existencia de ordinales y, por tanto, de la longitud de la iteración. La reformulación es ésta: si asignamos un ordinal a cada elemento de un conjunto, hay un ordinal mayor que todos los ordinales asignados[15]. En esta forma, el axioma se basa en la descripción cantoriana de la generación de ordinales, pues si de los ordinales asignados uno es el máximo, su sucesor inmediato es mayor que todos ellos, mientras que si ninguno es máximo, el ordinal en cuestión se obtiene por aplicación del segundo principio. El axioma de substitución articula parte de las intuiciones de Cantor sobre la infinitud absoluta de los ordinales.

Ésta es la lista completa de los axiomas de ZF. Hay, sin embargo, otro axioma, el axioma de *elección*, que suele separarse de los restantes, no por razones de aceptación, sino por su carácter no constructivo. Los axiomas de existencia que hemos considerado hasta ahora dicen que cierto conjunto, cuya definición el axioma propor-

14. Con más detalle, el axioma de infinitud dice que existe un conjunto a tal que $\varnothing \in a$ y, para todo $x \in a$, $x \cup \{x\} \in a$. Formularlo así evita tener que definir el concepto de infinitud antes de presentar el axioma. No es difícil ver que cualquier conjunto que cumpla estas dos condiciones debe ser infinito.

15. Esta formulación en términos de ordinales implica (según la descripción de la jerarquía acumulativa que hemos esbozado) la formulación original: puesto que cada conjunto aparece en algún estrato, sea V_{ax} el primer estrato en el que aparece a_x. Si β es un ordinal mayor que todos los a_x, entonces V_β es un conjunto al que todos los a_x pertenecen.

ciona, existe. El axioma de elección, por el contrario, afirma que existe un conjunto que cumple cierta condición, pero del cual no ofrece ninguna descripción: es un principio de existencia pura. Una de sus múltiples formulaciones equivalentes es ésta: Dado un conjunto *a* y una partición de *a* en conjuntos no vacíos y sin elementos en común, existe un subconjunto *b* de *a* que contiene exactamente un elemento de cada una de las partes de la partición.

La justificación de este axioma requiere un análisis más detallado del concepto de conjunto de objetos de un dominio. Más adelante nos ocuparemos con detalle de esta cuestión, de notable importancia filosófica. El punto crucial es decidir si todo conjunto está determinado por una propiedad. Naturalmente, el único modo de presentar un conjunto infinito es mediante una condición que sus elementos y sólo ellos cumplen, pero ello no significa que no existan conjuntos que no puedan ser presentados. Además de los conjuntos *definibles o especificables* puede haber conjuntos *combinatorios*, o sea, meras combinaciones arbitrarias de objetos de un dominio, selecciones aleatorias obtenidas sin seguir ninguna regla. Ésta es una idea que requiere elucidación, ya que no es claro su significado. En todo caso, la aceptación del axioma de elección está ligada a la adopción de esta idea combinatoria de conjunto. Si la aceptamos como coherente, el axioma de elección es fácilmente justificable: aunque no podamos dar ninguna condición que sea satisfecha por un elemento de cada parte y sólo uno, entre las múltiples combinaciones de elementos de *a* debe haber una que contenga exactamente un elemento de cada una de las partes. La idea de conjunto combinatorio y, con ella, la aceptación del axioma de elección son elementos constitutivos de la teoría de conjuntos actual procedente de Cantor. Para Russell, por el contrario, «toda clase está definida por alguna función proposicional que es verdadera de los miembros de la clase y falsa de lo demás» (Russell, 1919, 183); en consecuencia, aunque se vio obligado a usar el axioma de elección, hubo de reconocer la imposibilidad de justificarlo, ya que «a no ser que podamos hallar una *regla* para seleccionar [un objeto de cada parte], no sabremos que una selección es siquiera teóricamente posible» (Russell, 1919, 126).

La teoría de Zermelo Fraenkel con el axioma de elección se designa como ZFC («C» por *choice*, elección). Es ésta la teoría considerada como fundamento de la matemática en el sentido preciso que dimos a este término.

7. REALISMO METODOLÓGICO

Las ideas y los conceptos con que forjamos la concepción iterativa y que subyacen a la teoría axiomática no son formulables con precisión. Es el caso del concepto de conjunto combinatorio y de la idea de generabilidad absolutamente infinita de los ordinales. Por esta razón, estos conceptos e ideas no aparecen en los axiomas ni en los teoremas que de ellos se siguen; pero no son ociosos, ya que iluminan el trasfondo de la teoría matemática propiamente dicha. Además, a pesar de su imprecisión, nos han permitido justificar los axiomas de ZFC. Naturalmente, la justificación que hemos dado no es una demostración matemática; las demostraciones matemáticas en teoría de conjuntos parten precisamente de los axiomas.

En el paso de la concepción informal a la axiomatización se pierden algunos componentes de gran importancia motivacional. Así, los ordinales y, con ellos, los estratos de la jerarquía acumulativa, son concebidos como siempre en proceso de expansión; el universo de los conjuntos, visto desde la perspectiva preaxiomática cantoriana, es un universo abierto (ni los conjuntos ni los ordinales están todos presentes a la vez para formar un dominio), por lo que no debe sorprendernos que a Cantor no le afectara la paradoja de Russell y negara la validez del principio ilimitado de comprensión.

Pero en los axiomas no se aprecia nada de ello. Más aún, la lectura por así decir estática que hacemos de los cuantificadores, en particular del cuantificador existencial, nos induce a pensar, cuando estamos inmersos en la teoría propiamente matemática, que con cada teorema descubrimos o describimos un aspecto más o menos importante de un mundo cerrado y bien delimitado. Así, el segundo principio cantoriano de generación, apelando al cual Cantor motivaba la infinitud absoluta de los ordinales, se traduce en el siguiente teorema: «Todo conjunto de ordinales está acotado superiormente»; la cota no se crea: está ahí.

Algo semejante ocurre con respecto al concepto de conjunto combinatorio. Las dificultades inherentes a esta noción, que examinaremos más adelante, no pueden recogerse en los axiomas. En nuestra concepción informal de los conjuntos que está en la base de la teoría, distinguimos entre conjuntos definibles, que podemos construir o, por lo menos, especificar, y conjuntos meramente combinatorios, de los cuales somos incapaces de dar un solo ejemplo. En la teoría, esta diferencia se desvanece; no hay distintos grados de existencia.

Es importante, pues, distinguir la actitud del matemático de la del filósofo ante la teoría de conjuntos (en general ante cualquier teoría

matemática), incluso en el caso de que matemático y filósofo sean una misma persona. El uso que hacen uno y otro de la teoría es distinto, como son distintos los criterios de justificación de principios y de métodos[16]. La teoría de conjuntos es una teoría matemática y no es tarea del filósofo enmendarla. El filósofo debe respetar lo que podemos llamar «la letra» de teoría, o sea, sus axiomas, sus teoremas y sus procedimientos de demostración; el respeto a la letra es la única frontera a su labor interpretativa. Puede haber distintas interpretaciones coherentes mutuamente incompatibles de la misma teoría, y es en este ámbito de la interpretación donde actúa el filósofo.

Un aspecto característico de la matemática conjuntista es el uso generalizado de métodos de inferencia no constructivos y la admisión incondicional de demostraciones de existencia pura; se demuestra la existencia de conjuntos de los cuales no es posible dar definición alguna. Un famoso ejemplo de teorema de existencia pura es la llamada *paradoja de Banach-Tarski*, que podemos formular así: Dadas dos esferas cualesquiera *A y B*, existe una partición de cada una de ellas en un mismo número finito, n, de partes

$$A = A_1 \cup A_2\, U \cup \ldots \cup A_n \quad \text{y} \quad B = B_1 \cup B_2 \ldots \cup B_n$$

tales que A_1 es congruente con B_1, A_2 *lo* es con B_2, ..., y A_n lo es con B_n. Para poner de manifiesto su aspecto paradójico, este teorema puede reformularse diciendo que podemos dividir una esfera del tamaño de un guisante en un número finito de partes y reagruparlas usando sólo movimientos rígidos (traslaciones, rotaciones, reflexiones, etc.), para formar una esfera del tamaño del Sol. Esta reformulación es incorrecta en sólo un aspecto que pone en evidencia su contenido puramente existencial: *no podemos* dividir la esfera en estas partes, sino que la división de la esfera en estas partes *existe*. Si pudiéramos efectuar esta división, las distintas partes serían cuerpos cada uno de las cuales tendría un volumen determinado. Puesto que el volumen de un cuerpo se conserva al someterlo a movimientos rígidos, el volumen de la esfera original y el de la esfera final serían ambos la suma de los volúmenes de las partes: ambas tendrían el mismo tamaño. Las partes

16. A este respecto es instructiva la actitud de Cantor con respecto a los ordinales. Si bien los introdujo con ayuda de los principios de generación, más adelante prefirió definirlos por abstracción a partir de buenos órdenes. La nueva definición no le permitía justificar la infinitud absoluta de los ordinales, tan importante para su concepción de la teoría. No obstante, decidió abandonar la definición original para, en sus propias palabras, «tratar el asunto de modo puramente matemático» (carta de Cantor a Mittag-Lefier del 23 de septiembre de 1883, en Cantor, 1991, 130).

cuya existencia garantiza el teorema de Banach-Tarski son (por lo menos algunas) conjuntos sin volumen que no podemos construir. Pero existen, según la teoría.

Este ejemplo es una manifestación de lo que podemos llamar el realismo metodológico de la teoría de conjuntos, que se extiende a aquellas partes de la matemática que usan métodos conjuntistas de manera esencial. Lo característico del realismo metodológico es que se trata a los conjuntos y a los objetos matemáticos en general como si formaran parte de una realidad estática y bien determinada que los axiomas describen, por lo menos parcialmente. Otro índice del realismo metodológico es el uso de definiciones llamadas impredicativas, en las que caracterizamos un conjunto en términos de una colección de conjuntos entre los cuales se halla el conjunto mismo.

8. REALISMO FILOSÓFICO

El realismo metodológico no es una actitud filosófica, sino un modo de hacer matemática. Se caracteriza por los métodos de demostración, no sólo los ya citados del uso del axioma de elección y definiciones impredicativas, sino también otros más habituales, encarnados en la lógica clásica, como el principio del tercero excluido y las demostraciones por reducción al absurdo. El realismo metodológico es un dato del que la interpretación filosófica de la teoría debe dar cuenta, y el modo más directo de hacerlo es mediante el realismo filosófico, la aceptación de que, efectivamente, los conjuntos existen independientemente de nosotros, de nuestras definiciones y de nuestras construcciones, y el matemático los describe y los estudia. El realismo filosófico en matemáticas se conoce también como «platonismo»[17].

El realismo filosófico surge naturalmente si entendemos los enunciados matemáticos al pie de la letra y si suponemos que al demostrar un teorema justificamos su verdad. Tomado literalmente, un teorema de la teoría de conjuntos habla de conjuntos, expresa, por ejemplo, que ciertos conjuntos se hallan en determinadas relaciones mutuas, o que existe un conjunto que cumple cierta condición, etc. Para que el teorema sea verdadero debe haber conjuntos que se hallen en la relación descrita, o debe existir un conjunto que cumpla las condiciones en cuestión. La atribución de independencia de los conjuntos respec-

17. La oposición que hacemos entre realismo metodológico y realismo filosófico es análoga a la de Shapiro (1997) entre *working realism* y *philosophical realism* y a la de Ferreirós (1999b) entre platonismo interno y platonismo filosófico.

to a las construcciones y definiciones se apoya en la objetividad de la matemática y en la validez de los principios y procedimientos de demostración no constructivos propios del realismo metodológico.

Uno de los más conspicuos abogados del realismo filosófico en la matemática en general y en teoría de conjuntos en particular es Kurt Gödel. Su argumento en favor de la existencia de un reino de objetos matemáticos autónomos es que, sin ellos, no es posible dar cuenta de nuestras teorías matemáticas. «Las clases y los conceptos pueden concebirse como objetos reales», escribe en un famoso pasaje, y añade:

> La admisión de tales objetos me parece tan legítima como la suposición de los cuerpos físicos y hay las mismas razones para creer en su existencia. Son necesarios para obtener un sistema satisfactorio de las matemáticas en el mismo sentido en que los cuerpos físicos son necesarios para una teoría satisfactoria de nuestras percepciones sensoriales (Gödel, 1944, 128).

A pesar de su naturalidad, el realismo filosófico presenta serias dificultades, sobre todo de índole epistemológica. A diferencia de los objetos físicos con que estamos familiarizados, los objetos matemáticos cuya existencia admite el platonismo carecen de aquellas propiedades por las que podemos entrar en contacto con los objetos físicos, ni siquiera son entes espaciotemporales como nosotros. ¿Cómo podemos llegar a saber algo acerca de ellos?

Gödel habla de una facultad especial, la intuición matemática, que nos permite acceder al reino de los conjuntos:

> A pesar de su alejamiento de la experiencia sensorial, tenemos algo análogo a una percepción también de los objetos de la teoría de conjuntos, como se ve a partir del hecho de que los axiomas se nos imponen como verdaderos. No veo ninguna razón para tener menos confianza en esta clase de percepción, es decir, en la intuición matemática, que en la percepción sensorial, que nos induce a construir teorías físicas (Gödel, 1947, 268).

A diferencia de la kantiana, la intuición matemática postulada por Gödel es una facultad enigmática que pocos filósofos realistas están dispuestos a aceptar. Pero sin apelar a ella, ¿cómo explicar la existencia de conocimiento matemático, si éste versa sobre un mundo de objetos extrasensoriales? Un intento reciente de dar cuenta del conocimiento conjuntista como conocimiento sobre objetos lo ofrece Penelope Maddy en *Realism in Mathematics*, cuya estrategia se basa

en considerar conjuntos impuros[18] supuestamente ubicados en el universo espacio-temporal. En sus propias palabras, el filósofo realista «admite los conjuntos de objetos físicos en el mundo físico, asignándoles el lugar espacio-temporal que ocupa la materia de que están hechos sus elementos (y los elementos de sus elementos, etc.)» (Maddy, 1990, 178). Su localización espaciotemporal permite a Maddy elaborar una explicación de nuestro conocimiento sobre conjuntos impuros, e, indirectamente, de los conjuntos de que se ocupa ZFC, puesto que, observa Maddy, podemos suponer que «obtenemos conocimiento de los conjuntos puros por inferencia teórica a partir de nuestro conocimiento perceptual e intuitivo de los conjuntos impuros» (Maddy, 1990, 156).

Hay pocos intentos de explicar la teoría de conjuntos de un modo no realista, ya que las posturas no realistas más desarrolladas, que son distintas variedades del constructivismo, no dan cuenta de la totalidad de la matemática clásica y rechazan buena parte de la teoría de conjuntos[19]. Uno de ellos es el *estructuralismo modal*, sugerido por Putnam (1967) y elaborado y desarrollado por Hellman (1989). El estructuralismo es una concepción filosófica según la cual «el tema principal de la matemática no son los objetos matemáticos individuales sino las estructuras en que éstos se disponen. Los objetos de la matemática [...] son sólo átomos, puntos sin estructura, o posiciones en estructuras. Y como tales no tienen identidad ni caracteres distintivos fuera de una estructura» (Resnik, 1997, 201). El estructuralismo es compatible con el realismo (véase Shapiro, 1997, que contiene una discusión de las distintas variedades de estructuralismo), ya que si bien no se compromete con la existencia de objetos autónomos a los que hacen referencia los términos matemáticos singulares, admite, sin embargo, la existencia de estructuras, algunas de las cuales, en particular la estructura de que supuestamente se ocupa la teoría de conjuntos, son entidades más complejas que cualquier objeto propiamente matemático.

Según la versión no realista del estructuralismo, que ha sido calificada por su principal defensor como «estructuralismo sin estructuras» (Hellman, 1996), la matemática no se ocupa de estructuras existentes, sino sólo posibles; de ahí que se le conozca como *estructuralismo modal*. Un aspecto problemático de esta concepción tiene que ver con

18. Maddy se limita a considerar conjuntos cuya clausura transitiva contiene objetos físicos. Para la distinción entre conjuntos puros e impuros véase nota 9.

19. Para una presentación de distintas formas de constructivismo véase Troelstra y Van Dalen, 1988.

la caracterización de estructuras posibles particulares, lo cual requiere el uso de la lógica de segundo orden, que presupone que está determinado cuáles son todas las subcolecciones de una colección infinita de objetos (o más bien cuáles son todas las *posibles* subcolecciones de una *posible* colección infinita de objetos), una suposición nada elemental, como veremos. Otro es el concepto de posibilidad, que hay que tomar como primitivo, ya que la explicación habitual, en términos de los llamados «mundos posibles», no es aceptable en este contexto no realista.

9. ¿QUÉ ES UN CONJUNTO ARBITRARIO?

Hemos mencionado algunas concepciones filosóficas sobre cómo entender la teoría de conjuntos. El atractivo de las realistas reside en su aparente adecuación, pero los intentos de explicar el conocimiento acerca de los objetos y estructuras supuestamente existentes en el universo conjuntista no son muy convincentes. Estrictamente hablando, Gödel no explica nada, y Maddy no alcanza más allá de unos pocos principios muy simples, ya que el resto es extrapolación[20]. Para el realismo estructuralista, el conocimiento sobre una estructura se da en términos deductivos: en una estructura vale lo que se sigue de las condiciones que la caracterizan; sin embargo, subsiste el problema de por qué debe existir una estructura que satisfaga tales condiciones. Este problema no ha recibido solución satisfactoria[21]. En cuanto al estructuralismo modal, se basa en presupuestos de difícil justificación en un contexto no realista, como el uso esencial de la lógica de segundo orden y el peso dado a la noción de posibilidad, tomada como primitiva, una noción muy fuerte, ya que debe permitir que toda estructura de que se ocupa la matemática, por compleja que sea, sea posible[22].

En lo que sigue, esbozaremos otro modo de entender la teoría de conjuntos, también desde una perspectiva no realista. Veremos que es posible concebir la jerarquía acumulativa de un modo cuasi formal,

20. Maddy abandonó más adelante su realismo filosófico. Véase Maddy (1997).

21. Para S. Shapiro, que desarrolla con detalle una versión realista del estructuralismo, el principio fundamental de existencia es que toda teoría coherente describe una estructura. Pero no está claro qué sea una teoría coherente. Véase Shapiro, 1997, 95.

22. Otra reconstrucción no realista de la teoría de conjuntos se halla en Lavine (1994), donde se elabora con detalle una justificación de los axiomas de ZFC partiendo de la base de que el concepto de infinito es una extrapolación de la idea de indefinidamente grande, es decir, de una cantidad finita pero demasiado grande para ser contada.

pero en absoluto arbitrario. Si lo hacemos, aunque no nos comprometamos realmente con la existencia de los conjuntos, comprenderemos por qué aceptamos los principios conjuntistas que aceptamos. La idea es que la teoría de conjuntos no nos dice cómo son los conjuntos, sino cómo queremos que sean.

Puesto que los axiomas de ZFC son justificables a partir de la concepción iterativa, nos limitaremos a considerar la jerarquía acumulativa. Los dos parámetros de que depende la jerarquía son, como hemos visto, el paso de un dominio a su dominio potencia y la sucesión de los ordinales. El dominio potencia de un dominio *D* debe contener todos los conjuntos de elementos de *D*. Como veremos, no es claro qué sea un conjunto arbitrario de elementos de un dominio infinito; de hecho, nuestras ideas sobre conjuntos no nos permiten construir la supuesta totalidad de conjuntos de elementos de *D*, pero sabemos por qué necesitamos esta totalidad y sabemos articular lo que queremos de ella. No hace falta que admitamos seriamente su existencia; basta con que podamos razonar como si existiera. Como veremos, la descripción que daremos del conjunto potencia ayuda a comprender algunas dificultades de la teoría de conjuntos actual.

Acerca de cómo concebir la sucesión de los ordinales de modo no realista no diremos mucho, sobre todo porque el mismo modo como Cantor los introdujo en términos de reglas de generación es ya una descripción que puede entenderse de manera puramente formal o, si queremos, ideal. De la idea de generación es posible deducir las propiedades estructurales básicas de los ordinales, con independencia de que existan o no, y a partir de la idea de que la generación no ha de tener ningún límite concebible es posible extraer algunos principios que implican que los axiomas de ZF valen para los conjuntos ideales obtenidos por iteración de la operación potencia a lo largo de los ordinales.

Aclaremos nuestro punto de partida y nuestra meta. Partimos de un dominio *D* y queremos obtener su dominio potencia $\mathcal{P}(D)$, cuyos objetos son todos los conjuntos de objetos en *D*. Para mayor fluidez, llamaremos *D-objetos* a los objetos en *D* y *D-conjuntos* a los conjuntos de *D*-objetos.

La idea básica, que posteriormente deberemos corregir, es que un *D*-conjunto no es más que una colección, entendida como el resultado de un proceso de selección de *D*-objetos: los elegidos pertenecen a la colección, los no elegidos no pertenecen a ella. Esta noción de colección es perfectamente adecuada para dominios finitos, pues nos permite calcular, si *D* es finito, cuántas colecciones de *D*-objetos hay, y nos permite describirlas explícitamente en un orden preciso en tér-

minos de una enumeración de *D*. Esto significa que está determinado qué colecciones de D-objetos hay, por lo que podemos tratarlas como objetos de un dominio. En el caso finito, pues, las colecciones cumplen el papel de *D*-conjuntos y su totalidad es $\mathcal{P}(D)$.

Pero esta idea de selección es o bien demasiado limitada o bien demasiado imprecisa para el caso de dominios infinitos. Si *D* es infinito, todo proceso de selección deberá estar descrito por un regla o, más generalmente, por una condición que distinga a los elementos seleccionados de los no seleccionados. Pero entonces o bien fijamos con precisión los medios de selección admisibles o bien dejamos abierta la cuestión de qué medios admitimos. En el primer caso estará totalmente determinado qué colecciones hay, las colecciones formarán un dominio, pero este dominio será demasiado pobre para funcionar como el dominio potencia de *D* que la teoría de conjuntos requiere[23]. En el segundo caso, al rechazar cualquier limitación en los medios de selección, nos acercamos más a las exigencias de la teoría de conjuntos, pero no nos adecuamos plenamente a ella, ya que, en este caso, las colecciones no forman un dominio: no está determinado qué colecciones hay, porque no está determinado cuáles son los medios de selección posibles.

Podemos tratar de compatibilizar riqueza con determinación introduciendo la idea de una selección arbitraria, es decir, de una selección sin reglas, en analogía con el caso finito. Hablamos entonces de una colección combinatoria, el resultado de infinitos actos independientes de decisión, uno para cada *D*-objeto, aunque no sea posible especificar ninguna condición que cumplan los objetos seleccionados y sólo ellos. Se trata, naturalmente, de un modo de hablar metafórico, ya que ningún agente efectúa selección alguna.

Hay algo aparentemente incoherente en la idea de una colección combinatoria. Entendemos que la existencia de una colección no depende de ningún proceso particular de selección, ya que distintos procesos de selección pueden dar el mismo resultado, pero esta independencia de cada proceso de selección particular no es una independencia de todo proceso de selección. Dicho de otro modo, una colección seleccionada con medios arbitrarios no es lo mismo que una colección arbitraria, seleccionada sin medio alguno. Esto último no sabemos qué es. En palabras de Hermann Weyl:

> La idea de que un conjunto infinito es una «colección» reunida mediante infinitos actos individuales de selección [...] es absurda... Opon-

23. Además, la extensión del dominio de las colecciones será sensible a la riqueza de los medios aceptados.

go el concepto [predicativo, o sea definible] de conjunto y función [...] al concepto completamente vago de función que se ha hecho habitual en análisis desde Dirichlet y, con él, el concepto habitual de conjunto (Weyl, 1918, § 4).

Ésta es la dificultad en que nos encontramos: no sabemos precisar qué entendemos por colección combinatoria y, por tanto, no estamos legitimados a introducir $\mathcal{P}(D)$ como la totalidad de las colecciones combinatorias de D-objetos; de modo ligeramente distinto: no podemos definir «D-conjunto» como «colección combinatoria de D-objetos». Pero podemos tratar de invertir la situación: en vez de explicar primero qué es un D-conjunto arbitrario y luego definir $\mathcal{P}(D)$ como la totalidad de los D-conjuntos, podemos explicar primero qué es $\mathcal{P}(D)$ y luego definir D-conjunto como $\mathcal{P}(D)$-objeto[24].

Podemos tratar de describir formalmente $\mathcal{P}(D)$ porque sabemos para qué lo queremos. Queremos que sea un dominio en el que cada colección de objetos que pueda llegar a ser especificada con medios cualesquiera esté representada. No podemos construir un dominio tal. Pero podemos aislar las condiciones (algunas de ellas) que cumpliría si existiese, y pretender que existe un dominio que las cumple, con el compromiso de que todo cuanto supongamos acerca de él se siga de estas condiciones. Si mantenemos este compromiso, la pretensión de existencia se convierte en un mero recurso lingüístico[25]. Para la descripción formal de $\mathcal{P}(D)$ necesitamos tres términos: $\mathcal{P}(D)$, D, y la relación de pertenencia E (o más bien $\in_{\mathcal{P}(D)}$) entre D-objetos y D-conjuntos. La condición que exigimos es que $\mathcal{P}(D)$ sea un dominio *máximamente extensional* sobre D. Que $\mathcal{P}(D)$ sea extensional (con respecto a $\in$) significa que para cualesquiera dos D-conjuntos distintos x, y existe un D-objeto d tal que o bien $d \in x$ y $d \notin y$, o bien $d \in y$ y $d \notin x$. Que $\mathcal{P}(D)$ sea máximamente extensional significa que no es posible extenderlo sin violar la extensionalidad.

La idea es simple: quisiéramos que el dominio potencia de un dominio D constara de la totalidad de las colecciones posibles, es de-

24. Esto es, en cierto modo, lo que hace el realismo filosófico; desde una perspectiva realista no hace falta precisar qué es un conjunto combinatorio, basta con hacer ciertas sugerencias al respecto, pues, dado que los conjuntos combinatorios están ahí, si bien nuestras sugerencias son incapaces de caracterizarlos, pueden, no obstante, servir para identificarlos, separándolos de otras entidades. Pero como explicación de «conjunto combinatorio», la vía realista es poco eficaz, ya que presupone la existencia de $\mathcal{P}(D)$, pero no nos dice qué es, más allá de la «totalidad de los D-conjuntos».

25. Las dificultades de expresión en un contexto no realista se incrementan por la inadecuación del lenguaje, mucho más apto para tratar del mundo circundante. De ahí la necesidad de hablar a menudo en términos de ficción.

cir, de la totalidad de las colecciones seleccionadas con cualesquiera medios. Como hemos visto, estas colecciones no forman un dominio. Por tanto, no exigimos que los *D*-conjuntos sean colecciones. ¿Qué son? No nos importa. Por ello describimos $\mathcal{P}(D)$ de modo puramente formal o estructural, no decimos nada acerca de la naturaleza de sus objetos, los *D*-conjuntos. De todos modos, el requisito de extensionalidad nos permite verlos como colecciones: si x es un *D*-conjunto, podemos pensar en x como la colección de los *D*-objetos relacionados con x por $\in$, o sea, la colección de los *D*-objetos d tales que $d \in x$. Extensionalidad significa que *D*-conjuntos distintos corresponden a *D*-colecciones distintas[26].

La exigencia de maximalidad está motivada por el deseo de que en $\mathcal{P}(D)$ estén representadas todas las colecciones de *D*-objetos. La idea es que si C es una colección de *D*-objetos seleccionada con cualquier medio posible, haya un *D*-conjunto x tal que C sea la colección de los objetos $\in$-relacionados con x. Esto podemos exigirlo para aquellas colecciones para las que dispongamos de un medio de selección, pero ¿cómo garantizar que si, con medios ahora no disponibles, especificáramos una colección C de *D*-objetos, tal colección estaría representada en $\mathcal{P}(D)$? El requisito de maximalidad de $\mathcal{P}(D)$ cumple esta función, pues si C no estuviera representada en $\mathcal{P}(D)$, sería posible extender $\mathcal{P}(D)$ con ayuda de C respetando la extensionalidad; para ello bastaría añadir a $\mathcal{P}(D)$ un nuevo objeto x y ampliar $\in$ decretando que un *D*-objeto está relacionado con x si y sólo si este objeto está en C.

Si identificamos cada colección de *D*-objetos con el *D*-conjunto que le corresponde, podemos decir que $\mathcal{P}(D)$ consta de colecciones reales y de conjuntos ideales. Estos últimos cierran la pluralidad abierta de las colecciones, la sumergen en un dominio maximal. Pero esto es sólo una forma de expresión sugerente. Estrictamente hablando, sólo hay los *D*-objetos (nos han sido dados desde el principio) y las colecciones de *D*-objetos que sabemos especificar. $\mathcal{P}(D)$ no es más que un artificio introducido para estudiar con mayor fluidez las colecciones que sabemos especificar. Podemos explicar sin problemas la supuesta referencia a conjuntos ideales o ficticios: cuando parece que nos referimos a conjuntos inespecificados de *D*-objetos, sobre todo cuando cuantificamos sobre *D*-conjuntos (es decir, cuando decimos que hay un *D*-conjunto que cumple cierta condición, o que todos los

26. Hay que tomarse esta observación con un grano de sal, ya que para disponer de la colección en cuestión deberíamos disponer de x y de la relación $\in$, que sólo hemos introducido formalmente.

D-conjuntos tienen tal y cual propiedad), simplemente estamos razonando a partir del supuesto de un dominio máximamente extensional sobre *D*. Todas nuestras aseveraciones y conjeturas sobre *D*-conjuntos deben ser entendidas como basadas en la suposición de este dominio, sobre el cual podemos decir muchas cosas no arbitrarias gracias a nuestra comprensión de la idea de extensionalidad maximal. En particular, podemos obtener algunos axiomas básicos en los que basar nuestras investigaciones sobre conjuntos.

Esta explicación del dominio potencia de un dominio dado no es parte de la teoría de conjuntos, sino de su interpretación filosófica. Si la injertamos en la descripción esquemática de la jerarquía acumulativa que dimos más arriba, obtenemos una visión del universo conjuntista como algo cuasi formal. De cada uno de los estratos con índice finito sabemos dar una descripción exhaustiva, vemos efectivamente cada V_n como un dominio y los conjuntos en V_{n+1} como colecciones de objetos de V_n. También sabemos describir el contenido de V_ω, que no es más que la unión de los V_n. En el paso de V_ω a $V_{\omega+1}$ irrumpe la idealización, y se incrementa paulatinamente. Pero esto no es óbice para que razonemos sobre el contenido ficticio de los estratos transfinitos basándonos en la caracterización formal del dominio potencia como máximamente extensional. El concepto de maximalidad no es preciso, pero esta imprecisión no afecta al rigor matemático de la teoría de conjuntos, ya que este concepto no pertenece a ella, si bien está en la base conceptual en que se apoyan los axiomas.

Según nuestra interpretación del dominio potencia, no es estrictamente apropiado, en el caso de un dominio infinito *D*, hablar del contenido exacto de $\mathcal{P}(D)$. Según la forma en que *D* nos sea dado, podremos especificar algunas colecciones que corresponden a *D*-conjuntos, pero la supuesta riqueza de $\mathcal{P}(D)$ no se agota con ellas. Nuevamente, esto no se manifiesta en teoremas matemáticos, ya que en la teoría matemática tratamos a $\mathcal{P}(D)$ como un dominio. De todos modos, esta indeterminación real de $\mathcal{P}(D)$ puede ser la responsable de la enorme dificultad de decidir ciertas conjeturas que son independientes de ZFC[27] y que tratan de la existencia de ciertos subconjuntos. La más famosa de estas conjeturas es la *hipótesis del continuo* de Cantor, según la cual todo conjunto infinito de números reales es o bien numerable —o sea, biyectable con el conjunto de los números naturales— o bien de la cardinalidad del continuo —es decir, biyectable con el conjunto de los números reales—. Para ver que lo único que entra en juego es el conjunto potencia, podemos reformular la hipótesis de

27. Es decir, no son ni demostrables ni refutables a partir de los axiomas de ZFC.

Cantor en términos del contenido de los primeros estratos infinitos de la jerarquía acumulativa. La reformulación es ésta: todo subconjunto infinito en $V_{\omega+1}$ es biyectable con V_ω o con $V_{\omega+1}$. Puesto que toda biyección entre subconjuntos de $V_{\omega+1}$ puede representarse por un subconjunto de $V_{\omega+2}$, que la hipótesis del continuo valga o no depende de qué subconjuntos de $V_{\omega+2}$ haya, es decir, del contenido de $V_{\omega+3}$. Pero $V_{\omega+3}$ se obtiene a partir de V_ω, que contiene sólo conjuntos finitos y cuya estructura no es más compleja que la de los números naturales, mediante tres aplicaciones de la operación *potencia de*.

¿Tiene la hipótesis del continuo un valor de verdad determinado? Ésta, nuevamente, no es una pregunta matemática, sino filosófica, ya que, según el realismo metodológico propio de la práctica conjuntista, esta hipótesis, como cualquier proposición matemática sobre conjuntos, es verdadera o falsa, ya que esto está implícito en el uso de la lógica clásica. Si p es una proposición cualquiera, $p \vee \neg p$ es una tautología y, como tal, aceptada como verdadera; de acuerdo con el uso canónico del término «verdadero», una disyunción es verdadera si y sólo si uno de sus miembros lo es, y una negación es verdadera si y sólo si la proposición negada es falsa. De esto se sigue, si nos movemos en un plano puramente matemático y tomamos como p la hipótesis del continuo, que esta hipótesis es verdadera o es falsa.

Desde una perspectiva filosófica, la respuesta a la pregunta no es obvia. Del dominio potencia de un dominio dado D sólo sabemos que es máximamente extensional sobre D; pero no es que sólo sepamos esto, sino que esto es todo cuanto puede saberse de $\mathcal{P}(D)$, y no porque lo demás sea incognoscible, sino porque no hay realmente nada más. Lo fundamental de la cuestión es, pues, si nuestra idea de maximalidad es lo suficientemente rica para permitir dar una respuesta clara a la hipótesis del continuo. Como hemos dicho, la idea de maximalidad no es precisa, pero de la comprensión que de ella tenemos hemos podido extraer algunos principios precisos sobre qué D-conjuntos debe haber, principios que, como los axiomas de separación y de elección, están incorporados en ZFC. Si esos principios agotan nuestra idea de maximalidad, la respuesta es negativa: la hipótesis del continuo carece de valor de verdad. Pero no hay razón para mantener que nuestra idea de maximalidad se agote en su contribución a ZFC. Recordemos la motivación que hay tras ella: toda colección de D-objetos que sea posible seleccionar con cualesquiera medios debe estar representada en $\mathcal{P}(D)$. Cabe la esperanza (fundada en vista del gran esfuerzo que se dedica a esta cuestión en los últimos años) de que las investigaciones en teoría de conjuntos permitan descubrir un nuevo axioma que decida la conjetura, un axioma que, en el plano es-

trictamente matemático, hablará de la existencia de conjuntos, pero que, desde nuestra perspectiva filosófica, entenderemos como una manifestación precisa de qué conjuntos debe haber para asegurar la inviolabilidad de la maximalidad extensional del dominio potencia.

REFERENCIAS BIBLIOGRÁFICAS

Boolos, G. (1971), «The iterative conception of set»: *The Journal of Philosophy* 68, 215-232.

Cantor, G. (1932), *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, ed. por E. Zermelo, Springer, Berlin.

Cantor, G. (1991), *Briefe* (H. Meschkowski y W. Nilson [eds.], Springer, Berlin.

Dedekind, R. (1872), *Stetigkeit and irrationale Zahlen*, Vieweg, Brunswick. (Trad. esp. en Dedekind, 1998, 78-94.)

Dedekind, R. (1998), *¿Qué son y para qué sirven los números? y otros escritos sobre los fundamentos de la matemática*, ed. e introd. a cargo de J. Ferreirós, Alianza, Madrid.

Ferreirós, J. (1999), *Labyrinth of Thought*, Birkhäuser, Basel.

Ferreirós, J. (1999b), «Matemáticas y platonismo(s)»: *La Gaceta de la Real Sociedad Matemática Española*, vol. 2, n.º 3, 446-473.

Fraenkel, A., Bar-Hillel, Y. y Levy, A. (1973), *Foundations of set theory* (2.ª ed. rev.), North Holland, Amsterdam.

Hellman, G. (1989), *Mathematics without numbers*, Oxford University Press, Oxford.

Hellman, G. (1996), «Structuralism without structures»: *Philosophia Mathematica* (3), 4, 100-123.

Gödel, K. (1944), «Russell's mathematical logic», en Gödel, 1990, 119-141.

Gödel, K. (1947), «What is Cantor's continuum problem?», en Gödel, 1990, 176-187.

Gödel, K. (1990), *Collected Works. Volume II: Publications 1938-1974*, Oxford University Press, New York.

Grattan-Guinnes, I. (1971), «The correspondence between Cantor and Philip Jourdain»: *Jahresbericht der Deutschen Mathematiker-Vereinigung* 73, 111-130.

Lavine, S. (1994), *Understanding the infinite*, Harvard University Press, Cambridge, Massachusetts.

Levy, A. (1979), *Basic set theory*, Springer, Berlin.

Maddy. P. (1990), *Realism in mathematics*, Oxford University Press, Oxford.

Maddy. P. (1997), *Naturalism in mathematics*, Oxford University Press, Oxford.

Moschovachis, Y. (1994), *Notes on set theory*, Sringer, Berlin.

Potter, M. (1993), «Iterative set theory»: *The Philosophical Quarterly*, vol. 43, n.º 171, 178-193.

Putnam, H. (1967), «Mathematics without foundations»: *The Journal of Philosophy* 64, 5-22.

Quine, W. (1976), *The ways of paradox and other essays* (ed. rev. y ampl.), Harvard University Press, Cambridge, Massachusetts.

Rang, B. y Tomas, W. (1981), «Zermelo's discovery of Russell's paradox»: *Historia Mathematica*, 8, 15-22.

Resnik, M. (1977), *Mathematics as a science of patterns*, Oxford University Press, Oxford.

Russell, B. (1903), *The principles of Mathematics*, Allen and Unwin, London.

Russell, B. (1919), *Introduction to mathematical philosophy*, Allen and Unwin, London.

Shapiro, S. (1997), *Philosophy of mathematics. Structure and Ontology*, Oxford University Press, Oxford.

Scott, D. (1974), «Axiomatizing set theory», en *Axiomatic set theory*, en T. Jech (ed.), American Mathematical Society, Providence, 207-214.

Shoenfield, J. (1977), «Axioms of set theory», en J. Barwise (ed.), *Handbook of mathematical logic*, North Holland, Amsterdam, 321-344.

Troelstra A. y Van Dalen, D. (1988), *Constructivism in mathematics. An introduction*, North Holland, Amsterdam.

Vaught, R. ([2]1995), *Set theory. An Introduction*, Birkhäuser, Basel.

Weyl, H. (1918), *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*, Veit, Leipzig. (Trad. ingl. de S. Pollard y T. Bole: *The continuum. A critical examination of the foundation of analysis*, Dover, New York, 1994.)

Weyl, H. (1925), «Die heutige Erkenntnislage in der Mathematik», en K. Chandrasekharan (ed.), *Gesammelte Abhandlungen*, vol. 2, Springer, Berlin (1968), 511-542.

DIVERGENCIA Y RIVALIDAD ENTRE LÓGICAS

María Manzano

1. EL DEBATE

La comparación entre lógicas requiere una cierta caracterización de las mismas y un posicionamiento respecto a la siguiente cuestión:

¿Es la lógica clásica un sistema lógico universal, adecuado para ser utilizado en todas las áreas en las que se aplica la lógica?

Los argumentos a favor y en contra del uso de la lógica clásica frente a otras lógicas no clásicas son frecuentes en la historia de esta disciplina; e incluso el trazar fronteras y desterrar a una gran parte de los sistemas que extienden o se oponen a la de primer orden porque no son lógica fue durante años cosa corriente. La diversidad de lógicas existentes y la velocidad con la que se crean nuevas, unidas a la flexibilidad y amplitud de miras de los teóricos de la lógica actuales, hacen que hoy no se conciba una postura semejante. Es interesante a este respecto la lectura del libro que en 1994 editó Gabbay, en el que destacados especialistas se enfrentan a la pregunta.

1.1. *¿Qué es un sistema lógico?*

Obsérvese que ya no nos interrogamos «¿qué es "la" lógica?», sino que utilizamos el artículo indefinido.

¿Debemos empezar por definirlo? El concepto que se tome debería considerar, al menos en principio, la multiplicidad de lógicas que se usan en aplicaciones en inteligencia artificial (IA) y en *software*, así como las empleadas en filosofía, matemáticas y lingüística.

Como todos sabemos, el estudio de la lógica se remonta a los filósofos griegos[1] —aunque no se denominó así hasta cinco siglos más tarde—, que la concebían como el campo de investigación en el que se proporcionan los patrones o estrategias —*tópoi*— que pueden desarrollarse en el curso de una buena argumentación competitiva. Puesto que la mejor y más honesta manera de triunfar en una controversia es proporcionar una argumentación correcta, el interés práctico conduce al teórico, investigándose la inferencia válida. En los *Tópicos*, que se ocupan del razonamiento dialéctico, se recogen consideraciones no sólo lógicas, sino también psicológicas y lingüísticas. Dicho razonamiento se caracteriza por el uso de *entimemas* —esto es, silogismos cuyas premisas son sólo probables— a diferencia del razonamiento apodíctico, que toma premisas verdaderas y ciertas. En las *Categorías* se tipifican los predicados y los términos, creando así una ontología que se debate constantemente *entre las palabras y las cosas*.

Sin embargo, todos nosotros, supuestos seres racionales, empleamos la lógica cuando razonamos, asimilamos o procesamos la información que recibimos del entorno, cualquier tipo de información. (Somos lógicos porque somos seres humanos.)

En sentido coloquial se usa el adjetivo lógico no sólo para describir las reglas del razonamiento correcto, sino en una gran variedad de casos, más en consonancia con el sentido original del *lógos* de los griegos —relacionado con el lenguaje, la doctrina, la estructura del conocimiento, la razón, etc.—. Durante el siglo XX la lógica fue retornando a su extensión y diversificación originales estudiándose en ella no sólo el razonamiento matemático o filosófico, sino también otros fenómenos de gestión y transmisión de información, de toma de decisiones y de la acción y en general en casi todos los contextos gobernados por reglas.

1.1.1. Lógica dinámica

Hay un interés creciente por la estructura lógica de los procesos cognitivos que subyacen al razonamiento humano: la lógica tradicional se ocupaba del resultado de dichas acciones; esto es, del contenido del juicio —no de la acción—, del resultado del razonamiento —no del proceso argumentativo— y de las pruebas —no de su búsqueda—. Ahora también se estudian los mecanismos de la trasmisión de la información. A este respecto el libro de Johan van Benthem (1996), que

1. Aristóteles [322 a.C.], *Organon*. (Recopilación de sus enseñanzas, hecha a su muerte.)

investiga la lógica de la transmisión de la información, resulta especialmente fecundo.

En los años setenta (véase Fischer y Ladner, 1979) se creó la lógica dinámica de programas, que extendía a la modal de forma que fuera no sólo multimodal, sino que permitiera que las modalidades se combinasen para formar nuevas. Se inspiraba en ella haciendo que cada programa se comportase como una relación de accesibilidad sobre un universo de estados —que aquí representan a los estados de un ordenador; esto es, a sus registros de memoria—; la intuición es que un programa es un objeto dinámico capaz de transportar al ordenador de un estado a otro, de manera tal que refleje la forma en que como resultado de la ejecución de un programa cambia el valor de verdad de los enunciados a ellos asociados. Mediante composición, unión o clausura reflexiva y transitiva se generan programas nuevos. Los axiomas de cada lógica modal tratan de captar las propiedades de la relación de accesibilidad que les es propia —esto es, reflexiva en T, transitiva en 4, etc.— mientras que los de la lógica de programas expresan las propiedades de las operaciones entre programas.

Inspirada en la lógica dinámica de programas, nace la nueva lógica dinámica; aquí el paso de un estado a otro se produce mediante obtención de información.

1) *actuación de datos: up(P)*

$$\lambda xy \bullet x \subseteq y \wedge Py$$

2) *actuación minimal de datos: μ — up(P)*

$$\lambda xy \bullet x \subseteq y \wedge Py \wedge \neg\exists z(x \subset z \subset y \wedge Pz)$$

3) *descargar datos: down(P)*

$$\lambda xy \bullet x \supseteq y \wedge Py$$

4) *posibilidad: poss(P)*

$$\lambda x \bullet \exists y(x \subseteq y \wedge Py)$$

¿Hemos de incluir a esta nueva lógica dinámica?

¿No tendrá una arquitectura excesivamente compleja, difícil de vertebrar mediante una definición? ¿No corremos el riesgo de tener que dar una respuesta tan abstracta a la pregunta sobre la naturaleza de los sistemas lógicos que deje de ser operativa?

1.1.2. Lógica y teoría de juegos

Pero además, la argumentación puede entenderse como un *juego* cuyas reglas determinan la comunicación civilizada entre los individuos y la toma de decisiones. Los juegos formaron parte de la lógica en la Antigüedad. La idea intuitiva —presente en la *Retórica* de Aristóteles— es que la validez recoge las estrategias ganadoras en el debate. En lógica se dio más importancia a la teoría de la prueba y a la consecuencia semántica, siendo justamente en este último contexto donde se utilizaron juegos para construir y comparar modelos (véase Hodges, 1984). Sin embargo en los últimos años han aparecido estudios sobre juegos lógicos, relacionándolos con la teoría de juegos de la que se sirve la economía, y se está desarrollando una teoría general de la teoría de juegos[2].

¿Incluimos a la lógica de la teoría de juegos aquí?

1.1.3. Lógica con diagramas

Centrándonos en el razonamiento clásico de las matemáticas, habida cuenta de lo útiles que resultan los diagramas para la resolución de problemas varios, podemos investigar las reglas del razonamiento visual.

¿Hay una lógica visual, diagramática?

Ya Euler, Venn y Peirce advirtieron la importancia que tienen los diagramas en las pruebas matemáticas. Su valor didáctico no se puso nunca en duda: los profesores de lógica somos conscientes de lo útiles que son, por ejemplo, los diagramas de Venn en teoría de conjuntos y en lógica de predicados monarios; o de la diferencia en cuanto al aprendizaje se refiere entre los cálculos de árboles y los lineales. Algunos lógicos y matemáticos —entre ellos, Barwise y Etchemendy— han estado trabajando en un modelo de inferencia que sea capaz de incorporar una explicación del razonamiento válido a partir de representaciones no lingüísticas, creando así una lógica heterogénea en la cual se incluyen sistemas de representación diversos (véase Barwise y Etchemendy, 1994). En muchos casos los razonamientos diagramáticos no actúan como meras comparsas pues sustituyen con ventaja a los lingüísticos, probándose para ellos teoremas de completud y corrección. Ello requiere que su presentación como cálculo deductivo sea matemáticamente impecable, con el rigor, precisión y falta de ambigüedad que exigimos a cualquier cálculo lógico. Es interesante a este respecto Allwein y Barwise, 1996.

2. Me consta que Johan van Benthem está escribiendo un libro sobre este tema.

¿Hemos de incluir el razonamiento heterogéneo?, ¿el exclusivamente diagramático?

1.1.4. Programa de investigación

Si nos limitamos al estudio de los sistemas lógicos existentes, nuestra respuesta sobre la naturaleza de la lógica se torna contingente, haciéndose dependiente de los avatares históricos que produjeron como fruto los sistemas lógicos que tenemos hoy. Y puesto que la creación de otros nuevos no cesa, debemos realizar un corte temporal y nuestra respuesta será perecedera; pero ya sabemos que el dilema *diacronía/sincronía* se plantea en todos los ámbitos y que el historicismo no fue nunca popular entre los lógicos. Una posibilidad es la siguiente: en vez de estudiar los sistemas lógicos existentes para determinar dónde reside la «*logicidad*», podemos estudiarlos procurando desentrañar qué se intentaba resolver con ellos[3], confiando en que se halle en las preguntas planteadas, que no en las respuestas. Si se encontrase una *invariante lógica*, común a los casos en estudio, tal vez pudiéramos avanzar una respuesta: proponer unos *universales lógicos*. La solución obtenida por esta vía tendría la ventaja de abarcar más posibilidades y de señalar los huecos existentes; esto es, las necesidades que pudieran ser cubiertas mediante nuevos sistemas lógicos —por ejemplo, para el razonamiento diagramático así como para otros posibles razonamientos heterogéneos, abductivos, del diagnóstico y de la acción[4]—. Dejamos en suspenso este trabajo que incluye una definición conceptual de la lógica y pasamos a estudiar la variedad de lo que hay.

1.2. *La divergencia entre sistemas lógicos*

Por lo que respecta a la posibilidad de establecer una comparación entre lógicas, lo que hago a continuación es plantearlo en función de las que aparecen en la referencia[5] que es hoy comúnmente aceptada (Abramsky, Gabbay y Maibaum, 1992-2000; Gabay, Hogger y Robinson, 1993-1998, y Gabay y Guenthner, 2001).

3. Ésta es la solución propuesta en Alwein y Barwise, 1996.
4. Este proyecto está funcionando ya: *Summa Logicae en el siglo* XXI, véase a este respecto su biblioteca digital http://logicae.usal.es.
5. Soy consciente de que, gracias a Dov Gabbay, incluso esta referencia está «viva» en vigoroso proceso reproductivo.

1.2.1. Sistema lógico

La génesis de una lógica puede ser ésta: del universo matemático ϑ se seleccionan ciertas estructuras o modelos $\mathcal{A}$ de una cierta clase $\mathfrak{K}$, los que queremos estudiar. Para ello se introduce un lenguaje adecuado *L* y se define el concepto de verdad de una fórmula φ en un modelo, $\mathcal{A} \Vdash \varphi$. Con cada uno se asocia de forma natural el conjunto de todas las fórmulas verdaderas en él *Th(A)*, que tomadas en su conjunto describen con precisión a $\mathcal{A}$; son su exhaustiva caracterización con los recursos que brinda *L*. Todas las teorías de la clase de estructuras en estudio tienen un núcleo común, el de las fórmulas válidas en todas ellas —*VAL*— y que, consecuentemente, no describen a ninguna estructura en particular, describen a la lógica misma, son la lógica en estudio. Por consiguiente, podemos decir que la lógica es este conjunto *VAL*, que no es otra cosa que la gran intersección de todas las fórmulas en *Th(A)*, para cada $\mathcal{A} \in \mathfrak{R}$:

$$\cap \{Th(\mathcal{A}) \mid \mathcal{A} \in \mathfrak{R}\}$$

Frecuentemente, junto a la noción de verdad en un modelo se introduce la de consecuencia semántica de una fórmula a partir de un conjunto de ellas: usamos $\Gamma \Vdash_{\mathfrak{R}} \varphi$ para indicar que todos los modelos $A \in \mathfrak{R}$ de Γ lo son de φ. Esto es,

$$\forall \mathcal{A}(\mathcal{A} \in \mathfrak{R} \Rightarrow (\forall \gamma(\gamma \in \Gamma \Rightarrow \mathcal{A} \Vdash \gamma) \Rightarrow \mathcal{A} \Vdash \varphi))$$

El concepto semántico de consecuencia puede ser poco operativo, por lo que el paso siguiente suele ser el de intentar reemplazarlo por una réplica sintáctica, computable; esto es, mediante las reglas de un cálculo deductivo diseñadas para generar *VAL*. Decimos que φ es un teorema lógico —y escribimos $\vdash \varphi$— cuando dicha fórmula se ha conseguido siguiendo las reglas del cálculo. Por supuesto, el conjunto *THEO* de todos los teoremas lógicos y el conjunto *VAL* de las fórmulas válidas deben coincidir para que el objetivo de equivalencia entre sintaxis y semántica se haya cubierto, como veremos con detalle en el apartado 2.1.

Se ha descrito una lógica con una fuerte componente semántica, al menos en su génesis. Sin embargo, hay casos en los que el procedimiento se invierte y lo que tenemos en primera instancia es un cálculo deductivo: el concepto de consecuencia es sintáctico y la semántica se añade, si acaso, al final. Para permitir un acercamiento neutro al tema, en donde no primen los aspectos sintácticos sobre los semánti-

cos —ni a la inversa—, se puede definir un sistema lógico como un conjunto de fórmulas cerrado bajo consecuencia y fijar con precisión este concepto.

Como punto de partida Gabbay (1994) propone trabajar con una definición provisional, para la que es preciso concretar:

1. Un lenguaje formal *L*, incluyendo un alfabeto y una definición de fórmula.

2. La relación de consecuencia se establece entre conjuntos finitos de fórmulas, exigiéndole que satisfaga los requisitos de reflexividad, monotonía y corte. Puesto que la relación de consecuencia puede alcanzarse tanto sintácticamente como semánticamente —esto es, como deduciblidad $\vdash$ y como $\vDash$— en una definición no comprometida con ninguno de estos planteamientos, se habla de consecuencia como relación binaria entre conjuntos de fórmulas $\Delta \vdash \Gamma$[6].

3. Dicha relación puede definirse de varias maneras: se pueden listar los pares $\langle \Delta, \Gamma \rangle$ para los que $\Delta \vdash \Gamma$ se verifica, o incluso se pueden proporcionar los pares de conjuntos de fórmulas $\langle \Delta, \Gamma \rangle$ a un programa de un ordenador y esperar la respuesta.

4. Una semántica para *L* es una interpretación de las fórmulas de *L* en una familia de estructuras matemáticas que nos permita entender la relación de consecuencia $\vdash$ haciéndola depender de los modelos en los que sean verdaderas. Se puede definir un *modelo* como una función que a cada fórmula le asigna un valor en el conjunto $\{0,1\}$. Es decir,

$$Modelo\text{: } FORM(L) \rightarrow \{0,1\}$$

Una *semántica*, *S*, es un conjunto de modelos, junto a la definición de $\Delta \vdash \Gamma$ mediante la condición siguiente:

$$\forall s \in S\ [\forall X \in \Delta(s(X) = 1) \rightarrow \exists Y \in \Gamma(s(Y) = 1)]$$

5. Hay casos en los que para generar la relación $\vdash$ contamos con procedimientos algorítmicos, que pueden ser procedimientos de decisión o demostradores automáticos de teoremas.

6. El concepto de sistema o cálculo deductivo no está bien delimitado; tenemos ejemplos interesantes en sistemas de Gentzen, de Hilbert, *tableaux*, etc. Como primera aproximación podemos definir

6. Tarski define la consecuencia como una operación establecida sobre el conjunto de las partes, o conjunto potencia del de las fórmulas, y la caracteriza mediante una serie de axiomas.

un sistema deductivo como un procedimiento algorítmico para generar $\vdash$ que usa reglas de la forma

$$\frac{\Delta_1 \vdash \Gamma_1; \ldots; \Delta_n \vdash \Gamma_n}{\Delta \vdash \Gamma}$$

y axiomas de la forma

$$\frac{\varnothing}{\Delta \vdash \Gamma}$$

Los axiomas son los elementos iniciales del conjunto de pares $\langle \Delta, \Gamma \rangle$ que constituyen la relación de consecuencia, $\vdash$. Un cálculo deductivo $S_1 \vdash$ no es otra cosa que un modo de generar $\vdash$.

1.2.2. Propiedades matemáticas de $\vdash$

¿Qué propiedades matemáticas poseen dichas lógicas?

La relación clásica de consecuencia y la de los sistemas que la extienden cumple:

1. $\Delta \vdash A$, si $A \in \Delta$ *(reflexividad).*
2. $\Delta \vdash A \Rightarrow \Delta, \Delta' \vdash A$ *(monotonía).*
3. $\Delta \vdash B; \Delta, B \vdash A \Rightarrow \Delta \vdash A$ *(corte).*

Por el contrario, las lógicas no monotónicas violan esa ley y las subestructurales pueden carecer de la de corte. Por otra parte, muchas tienen una versión más estricta de reflexividad, $A \vdash A$. Estas propiedades constituyen un filtro matemático adecuado para juzgar y comparar lógicas[7]. Se tratará de una comparación de naturaleza estructural, pero interesante.

1.2.3. Equilibrio entre propiedades

Puesto que cada lógica intenta modelizar un determinado ámbito, cabe preguntarse hasta qué punto lo ha conseguido. Destacan, por un lado, la *naturalidad* del lenguaje, su adecuación como vehículo de expresión de las características de la parcela de *realidad* que está siendo

7. Hay quienes han querido situar las barricadas para defender las fronteras de la lógica en el cumplimiento de las tres propiedades, considerando que se podría admitir que hay otras relaciones interesantes entre proposiciones, pero que *no son lógica.*

estudiada, hasta qué punto los razonamientos se clarifican al ser formalizados. Por otro lado está la *eficiencia* computacional, si tenemos buenos cálculos, incluso si tenemos probadores automáticos que los implementen, automaticen y suplan. Una lógica es como una balanza: en un platillo se coloca el cálculo deductivo y se evalúan las *características computacionales*; en el otro, su *capacidad expresiva*.

¿Dónde está el fiel de la lógica que estudiamos?

Deberíamos comparar el uso de la lógica clásica con la no-clásica en sus aplicaciones a la IA y a otras áreas. Sabemos, por el teorema de Lindström, que la lógica de primer orden es la más equilibrada de una gran familia de lógicas.

1.2.4. Recursos matemáticos utilizados

Al definir un sistema lógico particular necesitamos utilizar una serie de recursos matemáticos, tanto al definir su cálculo deductivo, como al trazar su semántica. *¿Qué supuestos conceptuales hemos tenido que asumir?* Hay diferencias entre el uso de la lógica clásica y el de las no-clásicas. Si la primera pretende ser un sistema universal, ello debería notarse a este nivel: se tendría que asumir lo mínimo. Pondré un ejemplo: en la lógica de segundo orden estándar se puede expresar la hipótesis generalizada del continuo y el axioma de elección; ello nos obliga a decidir en qué entorno matemático nos movemos o, lo que es lo mismo, qué extensión de ZF estamos usando como metalenguaje. De manera que en esta lógica se asume la teoría de conjuntos a un nivel no intuitivo ni trivial[8].

1.2.5. Traducción

Para poder relacionar las diferentes lógicas, lo natural es construir puentes entre ellas; es decir, establecer correspondencias, traducciones, morfismos. Ciertamente toda deducción lógica puede ser simulada mediante una máquina de Turing. *¿Es suficiente?*

El concepto de traducción debe también ser clarificado. Mediante gödelización todas las lógicas pueden serlo, usando como lenguaje de llegada el de la aritmética; otra posibilidad es usar la teoría de conjuntos no sólo como semántica subyacente, sino como lenguaje de llegada; también sabemos que todo conjunto recursivamente enumerable puede generarse mediante cláusulas de Horn.

8. En Manzano (1996) utilizamos esta peculiaridad de la lógica de segundo orden estándar para demostrar su incompletud.

¿Nos basta con esta traducción?

Creo que en todos los casos anteriores priman las características del cálculo sobre la semántica de las fórmulas. Lo ideal es que junto a la traducción de las fórmulas se produzca una conversión bidireccional de las estructuras de la lógica en estudio a las de la lógica usada como referencia y que haya una correspondencia entre todos los componentes de los sistemas lógicos en estudio. Como veremos con detalle en la próxima sección, lo que propongo es convertir las estructuras en multivariadas, añadiendo universos en donde estén explícitamente representadas todas las categorías de objetos —conjuntos, relaciones, funciones, etc.— definibles en el lenguaje y usar el lenguaje y cálculo multivariado para realizar los cómputos.

1.2.6. Marco unificador

La elección de la lógica clásica como lógica unificadora está directamente relacionada con el éxito obtenido en la traducción de otras lógicas a ella; es decir, en su capacidad para servir como marco común.

¿Qué es un marco unificador?

Se podría usar la lógica clásica simplemente como metalenguaje formalizado e incluso implementar en el propio lenguaje objeto las características del metalenguaje. ¿Es esto deseable?

Hemos usado el término «*lógica clásica*»: ¿incluye a la heterogénea o multivariada?, ¿incluye a la de orden superior?

Mi objetivo es mostrar que la lógica clásica, si como tal incluimos a la multivariada o heterogénea, puede usarse como sistema universal, especialmente desde el punto de vista de la teoría de la prueba. Sin embargo, puesto que hay otros planteamientos globalizadores, expondremos dos de ellos para apreciar la conmensurabilidad de paradigmas.

No obstante, la existencia de un marco general en el que podamos representar los casos particulares no significa, necesariamente, que el mejor modo de tratar dichos casos sea inmersos en el general; simplemente da idea de la universalidad del tratamiento.

2. LÓGICA MULTIVARIADA COMO MARCO UNIFICADOR

2.1. *Introducción*

En *Extensions of first order logic* se estudian varias extensiones de la lógica de primer orden que tienen aplicación en filosofía, informática, matemáticas, lingüística e inteligencia artificial. En él se introdu-

cen con cierto detalle los siguientes sistemas lógicos: lógica de segundo orden (*SOL*), tres lenguajes para la teoría de tipos, lógicas modales proposicionales (*PML*) y lógica modal de primer orden-*S5*, lógica dinámica proposicional (*PDL*) y lógica multivariada o heterogénea (*MSL*). Una de las características del libro es que en él se administra una sustanciosa dosis de lo que Johan van Benthem denominaría *perspectiva lógica*. El primer objetivo es, sencillamente, el de presentar al lector una colección de lógicas —es relativamente detallada en el caso de la lógica de orden superior y de la multivariada, y como meros ejemplos de aplicación de la maquinaria de traducción propuesta para las modales y dinámica—; el segundo objetivo es desarrollar la tesis de que la mayoría de los sistemas *razonables* de lógica se traducen de forma natural a la lógica multivariada. Esta tesis se mantiene a lo largo de todo él, pero se desarrolla abierta y explícitamente en el último capítulo, en el que todas las lógicas consideradas se ponen en correspondencia directa con la lógica *MSL*. Hay otras muchas lógicas que pueden traducirse de forma semejante; el programa *traduccionista* de investigación, pese a llevar decenios funcionando, continúa operativo. No se me escapa que el esquema está pensado para lógicas que extienden a la clásica y que debería ser modificado oportunamente para otros casos, lo creo posible[9]. ¿Por qué *MSL*?, se podría preguntar.

MSL es una lógica muy flexible, resultando ser la elección obvia cuando se pretende tratar con más de un tipo de objetos. Pero además de ser la lógica *natural* para estos casos, es tambien una lógica *eficaz*, ya que su teoría de la prueba está bien desarrollada —posee un cálculo no sólo correcto, sino también completo—. Debido a su versatilidad, *MSL* puede ser usada como *marco unificador* en el que situar a otras lógicas.

Debido a la proliferación actual de lógicas es urgente y muy necesaria la reducción operativa de todas ellas. El objetivo es triple: 1) Usar en todos los casos un único cálculo deductivo y, a ser posible, un único demostrador de teoremas. 2) Arrastrar algunas de las metapropiedades desde *MSL* hasta otras lógicas en estudio. 3) Comparar entre sí a las diferentes lógicas, usando la comparación entre las teorías multivariadas que las representen.

Es difícil rastrear el origen y desarrollo de este enfoque, ya que cada lógica no clásica contó con su *doble* clásica casi desde su naci-

9. En las dos secciones posteriores analizaremos otros tratamientos unificadores que conducen por otras vías a Roma. Queda pendiente el aprovechar sus resultados para modificar nuestro esquema.

miento, existiendo incluso todo un campo de investigación dedicado a ello: la teoría de la correspondencia[10]. Sin embargo, debo[11] atribuir la mayoría de las ideas que yo uso en este planteamiento a dos artículos de Henkin que menciono en las referencias: «Completeness in the theory of types», de 1950, y «Banishing the rule of substitution for functional variables», de 1953.

Tampoco quisiera inducir a error. No hay en el artículo de 1950 traducciones de fórmulas, ni tan siquiera aparecen explícitamente el lenguaje y el cálculo multivariado. En conexión con la lógica de orden superior, el cálculo multivariado aparece por primera vez en el artículo de 1953. En él Henkin propone el axioma de definición de clases y relaciones —también llamado de comprensión— como alternativa a la regla de sustitución usada en los cálculos anteriormente propuestos para la lógica superior. Si de este nuevo se elimina el axioma de comprehensión, obtenemos un cálculo multivariado, es decir, el de *MSL*. Otra de las ventajas del que Henkin propone en 1953 es que permite aislar a otros cálculos situados entre el cálculo de *MSL* y los de orden superior. El procedimiento es sencillo: debilitando comprehensión. En el útimo capítulo esta idea se usa tanto en el estudio de la lógica dinámica como en el de la lógica modal.

2.1.1. «Completeness in the theory of types»

La prueba de completud del cálculo *MSL* de [17] está basada en la de Henkin de completud de la teoría de tipos. Aunque es una demostración conocida, pues es básicamente la misma que se usa en primer orden[12], se expondrá[13] brevemente a continuación para poder así señalar algunos aspectos que fueron decisivos en mi planteamiento.

El problema que nos ocupa es el siguiente: se nos presenta una determinada lógica, y además del cálculo de fórmulas, que nos dice qué filas de signos lo son, tenemos dos procedimentos para seleccionarlas: las reglas del cálculo deductivo, capaces de enumerar el conjunto de sus teoremas lógicos (*THEO*), y la semántica, que nos ayuda a escoger las fórmulas válidas (*VAL*).

10. Expuesta por Johan van Benthem (1986), quien es no sólo un *traduccionista* convencido, sino también el responsable de muchos de los resultados de la *teoría de correspondencia*.

11. Lo hago como tributo personal, con enorme placer.

12. Curiosamente, la demostración de Henkin de la completud para la teoría de tipos precede a la de completud de la lógica de primer orden.

13. He decidido centrarme en la lógica de segundo orden, que es un caso particular.

¿Coinciden estos conjuntos? Cuando $THEO \subseteq VAL$ decimos que el cálculo es correcto, y cuando $VAL \subseteq THEO$ decimos que es completo (en sentido débil). Desgraciadamente, para la lógica de segundo orden lo último no es cierto, ya que hay fórmulas válidas que no pueden derivarse en ningún cálculo. Sabemos, por el teorema de Gödel, que el conjunto de las fórmulas válidas en la semántica estándar es demasiado grande, inabarcable; dicho técnicamente, no es *recursivamente enumerable*. Por consiguiente, con la semántica estándar, el conjunto de las fórmulas válidas en *SOL* nunca podrá ser generado mediante un cálculo deductivo finito. La semántica estándar viene determinada por el uso exclusivo de estructuras

$$\mathcal{A} = \langle A, \langle A_n \rangle_{n \in N}, \ldots \rangle \text{ donde } A1 = \wp A,\ A_2 = \wp(A \times A), \text{ etcétera.}$$

Esto es, el concepto de *conjunto de elementos de A* que se usa para fijar A_1 es tomado de la metateoría de conjuntos, en el mismo sentido en el que en la lógica de primer orden (*FOL*) tomamos de la metateoría el concepto de identidad. Es decir, se toma como un concepto lógico o primitivo y no lo fijamos para cada estructura.

Pero incluso sabiendo que no hay cálculo deductivo capaz de seleccionar como teoremas todas las fórmulas válidas —*i.e.*, no existe ningún cálculo deductivo completo para *SOL*— podemos definir cálculos que sean correctos. Vamos a considerar dos de ellos: *MSL* es la extensión simple del cálculo para *FOL*, lo obtenemos al extender las reglas con cuantificadores para que den cuenta de las nuevas fórmulas, que son ahora fórmulas de segundo orden y tienen variables predicativas cuantificadas; el cálculo *SOL* lo obtenemos al añadir a *MSL* el esquema de comprehensión,

$$\exists X^n \forall x_1 \ldots x_n (X^n x_1 \ldots x_n \leftrightarrow \varphi) \text{ donde } X^n \text{ no libre en } \varphi.$$

Naturalmente ambos son incompletos, si tomamos como referencia la semántica estándar. Pero si dejamos abierto el concepto de *subconjunto de A* —esto es, *conjunto de elementos de A*—, es decir, si lo tomamos como un concepto a fijar para cada estructura, la situación cambia radicalmente. Esto fue justamente lo que hizo Henkin en 1950. La idea es muy sencilla: el conjunto de las fórmulas válidas es inabarcable porque el concepto de estructura estándar es demasiado restrictivo. Si aceptamos como estructuras adecuadas para *SOL* a las estructuras no-estándar —*i.e.*, estructuras donde $A_n \subseteq \wp A^n$, pero donde no se da necesariamente la igualdad—, el conjunto de las fórmulas válidas en ellas se reduce. De hecho, obtenemos un primer resultado de

completud para el cálculo *MSL* tomando esta semántica, que podemos denominar semántica de marcos. No obstante, si la selección de lo que vayamos a incluir en A_n es dejada completamente al arbitrio, el axioma de comprehensión podría fallar; es decir, algunos conjuntos definibles mediante fórmulas del lenguaje podrían haberse quedado fuera. Para que el cálculo *SOL* sea correcto, necesitamos la semántica de modelos generales de Henkin; esto es, los universos de la estructura deben estar cerrados bajo definibilidad. Podemos también demostrar la completud de *SOL*, bajo esta semántica, simplemente usando la completud de *MSL*. Por fortuna la propiedad de ser una estructura general puede ser axiomatizada mediante un conjunto de fórmulas Δ que incluye básicamente los axiomas de comprehensión. Esto es, tenemos

$$\models_{\mathfrak{G}.\mathfrak{S}} \varphi \Leftrightarrow \Delta \models_{\mathfrak{F}} \varphi$$

(donde $\mathfrak{F}$ es la clase de los marcos y $\mathfrak{G}.\mathfrak{S}$ la de las estructuras generales).

COMENTARIO 1. *Arrastrar metapropiedades de la lógica marco a la lógica en estudio es un recurso que usaremos continuamente; la clave para poderlo hacer es que sea axiomatizable en la lógica marco la propiedad de ser una estructura de la lógica en estudio.*

En realidad, al cambiar la semántica de modelos estándar por la de modelos generales, hemos cambiado de lógica.

COMENTARIO 2. *El planteamiento que se hace en esta prueba de completud de la lógica superior —completud respecto de modelos generales, naturalmente— es exportable. De ella aprendemos a modificar convenientemente la semántica de forma que las fórmulas válidas en la nueva semántica coincidan con los teoremas lógicos.*

La verdad es que no se trata, como cabría pensar, de una prueba *ad hoc*, ya que la nueva semántica es incluso más natural y razonable que la propia semántica estándar porque tomamos como *subconjunto de A* a los que son definibles en *A*. El resultado es que en la lógica de segundo orden estándar el fiel de la balanza se inclina del lado del poder expresivo, en detrimento de las cualidades lógicas, mientras que en la no-estándar está más equilibrada —se recuperan las propiedades lógicas de completud, compacidad y Löwenheim-Skolem.

Las ventajas de la nueva semántica son evidentes ya que: 1) conseguimos así un *teorema de completud*; 2) sirve para destacar no sólo

a la lógica *MSL* sino también a *otras lógicas* entre *MSL* y *SOL*, obtenidas *debilitando comprehensión*. La desventaja es que pierde parte de su potencia expresiva[14].

Aquí radica el programa de investigación que quiero comentar a continuación, pero antes indicaré dónde se sitúan esas lógicas intermedias.

2.1.2. «Banishing the rule of substitution...»

Hay otra idea, esta vez tomada del artículo de 1953, que también resulta interesante y que es la siguiente: Si debilitamos el axioma de comprehensión —por ejemplo, lo restringimos a fórmulas de primer orden, o a traducciones de fórmulas dinámicas o modales, o a cualquier otro conjunto recursivo— obtenemos cálculos *XL* entre *MSL* y *SOL*. Y resulta fácil encontrarles su semántica correspondiente. Naturalmente, la clase de estructuras que les corresponde $\mathfrak{S}\mathfrak{I}(XL)$ estará situada entre las anteriores, $\mathfrak{F}$ y $\mathfrak{G}.\mathfrak{S}$.

$$\mathfrak{G}.\mathfrak{S} \subseteq \mathfrak{S}\mathfrak{I}(XL) \subseteq \mathfrak{F}$$

La nueva lógica, llamémosla *XL*, será también completa —al menos en sentido débil—. El motivo es que esta clase de modelos es axiomatizable. Esta idea se explotará tanto en *PML* como en *PDL*, aunque con distinta fortuna.

2.2. *Reducción de otras lógicas a la multivariada*

Como dije anteriormente, en el último capítulo (Manzano, 1996) convierto otras lógicas a la multivariada, *MSL*.

Lo que haré ahora es explicar brevemente la idea general del programa de reducción.

2.2.1. Primer nivel: Teoremas de representación

El plan general es como sigue:

La signatura de la lógica en estudio —llamémosla *XL*— se transforma en una signatura multivariada, las expresiones de la lógica *XL* se traducen a *MSL* y las estructuras propias de *XL* se convierten en es-

14. En otro lugar (Manzano, 1998) comparo a la lógica de segundo orden no estándar con la *sirenita* del cuento de Andersen: obtiene con enorme sufrimiento unas piernas, e incluso un alma inmortal, pero debe entregar como precio su preciosa voz.

tructuras multivariadas. Por consiguiente, necesitamos definir una función recursiva *TRANS* para la traducción y una conversión directa de estructuras, $CONV_1$.

Una posibilidad es añadir a las estructuras multivariadas universos que contengan todas las categorías de objetos matemáticos de los que queramos —y podamos— hablar en la lógica *XL*. Por consiguiente, en $CONV_1(A)$ añadiremos universos que contengan los conjuntos y relaciones definibles en la estructura original *A* mediante los constructos de *XL*. A consecuencia de este cambio, parece que estemos tomando estructuras propias para *SOL* en vez de para *MSL*. Pero sabemos —lo aprendimos en la prueba de completud de Henkin— cómo evitar el uso explícito de *SOL*, cambiando a *MSL*: nuestros universos serán multivariadas o heterogéneos, pero añadiremos relaciones de pertenecia y el axioma de extensionalidad.

Con la conversión directa de estructuras queremos obtener como resultado la equivalencia entre la validez de una fórmula cualquiera φ en las estructuras originales de la lógica en estudio —llamémoslas simplemente $\mathfrak{S}\mathfrak{I}$ (*XL*)— y la validez de la clausura universal de su traducción a lógica multivariada $\forall TRANS(\varphi)$ en la clase $\mathfrak{S}^*$ de las estructuras obtenidas por conversión —donde $\mathfrak{S}^* = CONV_1(\mathfrak{S}\mathfrak{I}(XL))$—. Esto es,

$$\models_{\mathfrak{S}\mathfrak{I}(XL)} \varphi \Leftrightarrow \models_{\mathfrak{S}^*} \forall TRANS(\varphi)$$

Naturalmente, la primera pregunta que uno debe hacerse es si $\mathfrak{S}^*$ puede reemplazarse por la clase de modelos de un conjunto de fórmulas multivariadas a determinar, Δ. Por consiguiente, la clave para las definiciones de *TRANS* y de $CONV_1$ es el poder simplificar la demostración de la equivalencia semántica, y en este sentido la relevancia de los resultados que pudieran derivarse depende de si se puede o no axiomatizar $\mathfrak{S}^*$. En algunos casos, bastará con axiomatizar una clase algo más amplia, que incluya a $\mathfrak{S}^*$.

Resumiendo, nuestro primer objetivo es demostrar un teorema de representación; es decir, un teorema de esta forma:

$$\models_{\mathfrak{S}\mathfrak{I}(XL)} \varphi \Leftrightarrow \models_{\mathfrak{S}\mathfrak{I}(MSL)} \forall TRANS(\varphi)$$

A partir de este teorema se demuestra fácilmente el *teorema de enumerabilidad para la lógica XL*. Por lo tanto, sabemos que podremos diseñar un cálculo para *XL*, pero también sabemos que en *MSL* podemos simular ese cálculo e incluso *usar un demostrador de teoremas de MSL.*

Así que, aunque la reducción se hubiera detenido en este nivel, los resultados desde un punto de vista práctico serían impresionantes.

2.2.2. Segundo nivel: el teorema principal

Una vez demostrado el teorema de representación, lo inmediato es preguntarse:

¿Se puede hacer algo más?

¿Es el teorema de representación el objetivo último?

Cuando la lógica en estudio *XL* posee un concepto de consecuencia, podemos intentar demostrar *el teorema principal*; esto es, que consecuencia en *XL* equivale a consecuencia de las traducciones módulo la teoría Δ. Concretamente,

$$\Pi \vDash_{\mathfrak{S}\mathfrak{J}(XL)} \varphi \Leftrightarrow TRANS(\Pi) \cup \Delta \vDash_{\mathfrak{S}\mathfrak{J}(MSL)} TRANS(\varphi)$$

A partir del teorema principal se obtienen *compacidad y Löwenheim-Skolem para XL*, usando las correspondientes propiedades de *MSL*. Para demostrar el teorema principal hace falta introducir una conversión inversa de estructuras; nuestro objetivo al definirlas es demostrar que si partimos de un modelo de Δ, digamos $\mathcal{B}$, y de una fórmula de la lógica *XL* —$\varphi \in FORM(XL)$—, $\mathcal{B}$ es modelo de la clausura universal de la traducción de φ, $\overline{\forall}TRANS(\varphi)$, si y sólo si la conversión inversa de $\mathcal{B}$ lo es de φ. Esto es,

$$CONV_2(\mathcal{B}) \Vdash \overline{\forall}TRANS(\varphi)$$

2.2.3. Tercer nivel: equivalencia de los cálculos

Cuando la lógica en estudio *XL* poseee también un cálculo deductivo, podemos intentar utilizar la maquinaria de la correspondencia ya desarrollada para demostrar, si es posible, corrección y completud para *XL*.

Puesto que contamos ya con el teorema principal y con completud y corrección para *MSL*, nos gustaría completar el cuadro demostrando la doble flecha en la línea de abajo:

$$\begin{array}{c} TRANS(\Pi) \cup \Delta \Vdash_{\mathfrak{S}\mathfrak{J}(MSL)} TRANS(\varphi) \Leftrightarrow \Pi \Vdash_{\mathfrak{S}\mathfrak{J}(XL)} \varphi \\ \Updownarrow \\ TRANS(\Pi) \cup \Delta \vdash TRANS\ (\varphi) \Leftrightarrow \Pi \vdash \varphi \end{array}$$

La flecha hacia la izquierda es normalmente de demostración muy sencilla, ya que podemos añadir a Δ las condiciones requeridas, como

axiomas. En algunos casos resulta útil poder usar la construcción del modelo canónico.

2.3. *El ejemplo de la lógicas modales* K *y* S4

En esta sección lo que haremos es analizar el caso de la lógica modal normal —la que se interpreta en modelos de Kripke— y muy especialmente la minimal, *K y* la lógica *S4*. La traducción de fórmulas será la estándar, la conversión de estructuras consiste en añadir un universo en donde situar los conjuntos de mundos significativos y definiremos una teoría multivariada que representa y equivale a la modal, tanto en el caso de *K* como de *S4*.

2.3.1. El teorema de representación en PML

En este nivel se reducirá *PML* a $MSL^{\Box}$. Este lenguaje multivariado contiene los símbolos siguientes: $\bot$, $\neg$, $\rightarrow$, los relatores monarios P_0, P_1, P_2, ..., el relator binario *S*, un signo de pertenencia y dos de igualdad, para individuos y relaciones. Hay variables de dos tipos, para individuos y para conjuntos. Usaremos la conocida traducción estándar, que expresa en primer orden las condiciones de verdad de las fórmulas modales. En particular:

$$TRANS^{\Box}(\bot)[u] = u \neq u$$
$$TRANS^{\Box}(p)[u] = Pu$$
$$TRANS^{\Box}(\neg\varphi)[u] = \neg\, TRANS^{\Box}(\varphi)[u]$$
$$TRANS^{\Box}(\varphi \rightarrow \Psi)[u] = TRANS^{\Box}(\varphi)[u] \rightarrow TRANS^{\Box}(\Psi)[u])$$
$$TRANS^{\Box}(\Box\varphi)[u] = \forall v(Ruv \rightarrow TRANS^{\Box}(\varphi[v])$$

Es fácil ver que las traducciones de los axiomas *K y D f* $\Diamond$ expresan propiedades obvias de la cuantificación y que por lo tanto estas fórmulas pueden demostrarse en *MSL* como teoremas lógicos. Por otra parte, las traducciones tanto del axioma *T* como del axioma 4 no son fórmulas válidas en *MSL*, pues necesitan que se verifiquen ciertas hipótesis. ¿Cuáles?, nos preguntamos.

Las estructuras multivariadas que usaremos se obtienen a partir de las estructuras modales al añadir universos que contengan conjuntos de estados, o mundos. La intuición aquí es que queremos tener explícitamente representados todos los conjuntos que pueden ser definidos modalmente —mediante fórmulas modales en estructuras de Kripke.

Dada una estructura modal *A* construimos un marco al añadir un universo de subconjuntos de *W*, *W'*. Una clase muy especial de mar-

cos son los modelos generales construidos sobre estructuras modales. Para construir un modelo general de esta clase lo que hacemos es incluir en *W′* el conjunto *DEF* de los subconjuntos de *W* definibles algebraicamente —esto es, $DEF \subseteq W'$—. *DEF* contiene el conjunto vacío $\varnothing$ todo *W*, y las interpretaciones de las proposiciones atómicas. *DEF* está cerrado bajo operaciones booleanas; esto es,

$$\forall T\ (T, S \in DEF \Rightarrow T \cup S,\ W - T \in DEF)$$

y bajo la siguiente operación

$$\forall T(T \in DEF \Rightarrow Dom(R \cap (W \times T)) \in DEF)$$

Fuera de *DEF*, pero en *W′*, tenemos los unitarios de todos los elementos de *W*. La pertenencia a *W′* también obedece esta regla:

$$\forall w(w \in W \Rightarrow Rec(R \cap (\{w\} \times W)) \in W')$$

Llamemos $\mathfrak{F}$ a la clase de todos los *marcos construidos sobre estructuras modales* y $\mathfrak{G}$ a la clase de todas las *estructuras generales construidas sobre estructuras modales.*

Se puede demostrar que $\mathfrak{G}$ está contenido en la clase de marcos construidos sobre estructuras modales cuyo universo *W′* contiene todos los conjuntos definibles mediante fórmulas modales en sus propias estructuras modales. Asímismo, se puede demostrar que una fórmula modal φ define en una estructura modal $\mathcal{A}$ el mismo conjunto que su traducción define en cualquier estructura general construida sobre $\mathcal{A}$; esto es, $\mathcal{A}\mathfrak{G}$. Es decir,

$$\mathcal{A}\mathfrak{G}(\lambda u TRANS(\varphi)[u]) = \mathcal{A}(\varphi)$$

(Aquí $\mathcal{A}(\varphi)$ es el conjunto de los mundos de $\mathcal{A}$ donde la fórmula φ es verdadera.)

Finalmente, demostramos que la validez de una fórmula φ en *PML* equivale a la validez de la clausura universal de su traducción en la clase $\mathfrak{G}$. Es decir,

$$\vDash \varphi \text{ en } PML \Leftrightarrow \vDash_{\mathfrak{G}} \forall u TRANS(\varphi)[u]$$

Ahora la pregunta es: ¿podemos axiomatizar $\mathfrak{G}$?

La respuesta en este caso es afirmativa, ya que definimos la teoría *MODO* cuyos axiomas son los esquemas de comprehensión para tra-

ducciones de fórmulas modales, para fórmulas atómicas con igualdad y para átomos que usan el relator binario S —que representa a la relación de accesibilidad—. Esto es,

$$\forall\exists X\forall u(\varepsilon uX \leftrightarrow \varphi) \text{ donde } \varphi \in TRANS(PML) \cup I \cup \Sigma$$

I contiene las fórmulas de la forma $u = v$, y Σ contiene todas las fórmulas de la forma Suv.

Añadiremos extensionalidad, porque queremos simular la lógica de segundo orden en la multivariada.

Se demuestra lo siguiente:

$$Mod(MODO) = \mathfrak{G}$$

Usando estos resultados obtenemos un *teorema de representación* para la lógica minimal K:

$$\vDash \varphi \text{ en } PML \Leftrightarrow MODO \vDash_{\mathfrak{F}} \forall u TRANS(\varphi)[u] \text{ en } MSL$$

A partir del teorema de representación se demuestra *enumerabilidad* para la lógica K de *PML*. Además cualquier probador de teoremas para *MSL* podrá ser usado para demostrar teoremas modales.

Ya que tenemos un teorema de representación para K, buscamos uno para la lógica *S4*. Llamemos *MODO(S4)* a la teoría multivariada obtenida al añadir a *MODO* las condiciones abstractas de segundo orden correspondientes a las fórmulas modales T y *4* —llamémoslas *MS(T) y MS(4)*—. Sea $\mathfrak{D}$ la clase de las estructuras de Kripke reflexivas y transitivas. El teorema de representación de *S4* tiene esta forma:

$$\vDash_{\mathfrak{D}} \varphi \text{ en } PML \Leftrightarrow MODO(S4) \vDash_{\mathfrak{F}} \forall u TRANS(\varphi)[u] \text{ en } MSL$$

Este resultado es de fácil obtención porque en el cálculo multivariado se puede demostrar que las formulaciones habituales de reflexividad y transitividad equivalen a *MS(T) y MS(4)* módulo la teoría *MODO*, obteniendo de esta manera el siguiente resultado crucial:

$$MODO \vdash MS(T) \leftrightarrow Reflexividad$$
$$y\ MODO \vdash MS(4) \leftrightarrow Transitividad$$

COMENTARIO 3. *Esta situación es mejor que la que encontramos en PML con la semántica de modelos. Allí una estructura de Kripke reflexiva es un modelo de T, una transitiva lo es de 4, etc., pero hay mo-*

delos irreflexivos de T y modelos intransitivos de 4. Por lo tanto, si pensamos que los axiomas modales tratan de definir la relación de accesibilidad que les es propia, esto puede ser considerado como un cierto fracaso. Verdad es que se puede tomar la conocida como semántica de marcos para PML. La alternativa que yo propongo es que vayamos al entorno de las estructuras generales aquí definidas, puesto que nos ofrece una caracterización de las propiedades semánticas de la relación de accesibilidad sin perder el cálculo.

2.3.2. El teorema principal en PML

Puesto que ya tenemos el teorema de representación y en lógica modal contamos con el concepto de consecuencia, intentamos conseguir el teorema principal. En *PML* la conversión inversa es la función inversa de $CONV_1$ —borramos el universo *W'*— y así obtenemos fácilmente el *teorema principal* tanto para *K* como para *S4*. Esto es,

$$\Pi \vDash \varphi \Leftrightarrow TRANS(\Pi) \cup MODO \vDash TRANS(\varphi)$$

y

$$\Pi \vDash_{\mathfrak{D}} \varphi \Leftrightarrow TRANS(\Pi) \cup MODO(S4) \vDash TRANS(\varphi)$$

siendo $\mathfrak{D}$ la clase de estructuras de Kripke reflexivas y transitivas.

Como se mencionó antes, así conseguimos gratis los teoremas de *compacidad* y de *Löwenheim-Skolem* para las lógicas *K* y *S4*. La demostración se basa en el siguiente resultado crucial (5) arriba reseñado.

2.3.3. Los cálculos modales K y S4

Queremos demostrar que los cálculos modales para *K* y *S4* tienen un correlato en *MSL*. Es decir, que las teorías multivariadas que las representan, *MODO* y *MODO(S4)*, son también equivalentes sintácticamente a *K* y *S4*. Es decir,

$$TRANS(\Pi) \cup MODO \vdash TRANS(\varphi) \Leftrightarrow \Pi \vdash_K \varphi$$

y

$$TRANS(\Pi) \cup MODO(S4) \vdash TRANS(\varphi) \Leftrightarrow \Pi \vdash_{S4} \varphi$$

Por lo que a la lógica *K* se refiere, la flecha hacia la izquierda se obtiene fácilmente, porque las traducciones tanto de *K* como de *Df*◊

expresan propiedades obvias de la cuantificación y la regla de necesariedad se preserva bajo traducción. Por lo que respecta a la lógica *S4*, usando *MS(T)*, *MS(4)* y comprehensión obtenemos la traducción deseada de cada ocurrencia de *T* o de *4*.

En lógica modal, dada una lógica cualquiera, digamos *B*, hay un modelo canónico, $\mathcal{B}_B$, cuyo universo *W* contiene todos los conjuntos *B-máximamente consistentes* y cuya relación de accesibilidad se define así:

$$\{\langle x, y\rangle \mid \{\varphi \mid \Box\varphi \in x\} \subseteq y\}$$

A partir de este modelo, mediante conversión directa, obtenemos el modelo canónico general, $\mathcal{B}_B\mathfrak{G}$.

En el caso de la lógica *K* esta estructura no es tan sólo un modelo de *MODO*, sino que también podemos demostrar que la traducción de una fórmula modal φ es verdadera en un punto del modelo canónico si y sólo si la fórmula pertenece a dicho punto:

$$\mathcal{B}_K\mathfrak{G}[t] \Vdash TRANS(\varphi)[u] \Leftrightarrow \varphi \in t$$

Ahora podemos demostrar que

$$\mathcal{B}_K\mathfrak{G} \Vdash \forall u TRANS(\varphi)[u] \Rightarrow \vdash_K \varphi$$

Concluimos finalmente que

$$MODO \vdash_{MSL} \forall u TRANS(\varphi)[u] \Leftrightarrow \vdash_K \varphi$$

Usando este resultado llegamos a la completud y corrección de la lógica minimal de *PML*, *K*. No hemos hecho una demostración directa, la hemos traído de *MSL*.

Para la lógica *S4* se obtiene un resultado similar:

$$MODO(S4) \vdash_{MSL} \forall u TRANS(\varphi)[u] \Leftrightarrow \vdash_{S4} \varphi$$

Ahora ya tenemos los teoremas de *completud y corrección* para *K* y *S4*.

COMENTARIO 4. *Si alguien me preguntara qué hemos obtenido, la respuesta sería que ahora tenemos una serie de lógicas atrapadas en nuestra red multivariada. Creo que pescaremos a otras, aunque puede que tengamos que cambiar «las artes».*

2.4. *Lógica multivariada y parcial: un paso más*

Una situación que se presenta con cierta frecuencia es la siguiente: tenemos un sistema lógico no totalmente desarrollado o aceptado —al que llamaremos aquí la *Prelógica XL*— que se ha definido de manera puramente sintáctica —esto es, como un conjunto de fórmulas—, pero su semántica tal vez no esté fijada, o contamos tan sólo con una clase reducida de interpretaciones o modelos posibles. Se propone entonces la construcción de una lógica multivariada y posiblemente también parcial, *PHL*, como procedimiento para estudiar y mejorar la lógica *XL*. La nueva servirá de lógica subyacente sobre la que definiremos, mediante un proceso de traducción, la teoría que representa a la lógica *XL*.

La nueva lógica parcial y multivariada será lo suficientemente potente como para axiomatizar en ella las características semánticas de la lógica *XL*; contaremos además con un cálculo deductivo tipo Gentzen para demostrarlas.

Para construir la lógica marco *PHL* empezamos elaborando un cuestionario para *XL*, cuyos ítems dependen de la propia lógica en estudio, pero que están basados en estas dos preguntas:

1. Los modelos o interpretaciones estándar —o las simplemente buscadas, deseadas o pretendidas— ¿son extensionales o intensionales? En el primer caso, ¿cuántos tipos o variedades de objetos tenemos?

2. ¿Es un sistema bivariado o hay fórmulas que no reciben valor verdadero ni falso?

El caso que desarrolla con detalle Antonia Huertas (1994) —el de la lógica modal de primer orden— responde afirmativamente a ambas preguntas, así que es preciso construir una lógica parcial y multivariada: hay que definir su semántica y su sintaxis de manera que las propiedades de *XL* que queremos estudiar y expresar se puedan formular explícitamente en *PHL*.

En primer lugar hay que encontrar un conjunto de conectivas funcionalmente completo para la nueva lógica multivariada y parcial, al que exigimos que sea además *significativo*. Hay que definir incluso la propia relación de consecuencia; elegir entre consecuencia local y global, fuerte o débil, etc.[15]. Esta búsqueda requiere un estudio detallado bidireccional en el que intervienen los factores siguientes: la naturaleza de la parcialidad que queremos caracterizar; la adecuación del concepto de consecuencia semántica de la nueva lógica *PHL*; si

15. Algunas de estas posibilidades aparecen porque «no verdadero» no es aquí lo mismo que «falso».

queremos que conserve en lo posible las reglas clásicas de conectores, y si deseamos tener cuantificadores clásicos o no. Otro asunto verdaderamente complicado cuando tratamos con valores de verdad no clásicos es el de la introducción de la igualdad.

En segundo lugar presentamos un cálculo deductivo tipo Gentzen para *PHL*, definiendo reglas de introducción y eliminación de conectivas y reglas para los cuantificadores. (Es preferible este tipo de cálculo porque, al tener la propiedad de la subfórmula, permite ir hacia atrás en las deducciones y acoplar el cálculo de *PHL* al de la la nueva lógica que resultará al final de todo el proceso; esto es, la lógica *PXL* de la que hablaremos luego.)

Construido el marco *PHL*, inspirado en el cuestionario inicial, lo que hacemos es traducir la lógica en estudio, *XL*, a esta nueva lógica subyacente; para ello se traducen su signatura y sus fórmulas y se convierten las estructuras de la lógica *XL* en estructuras propias de *PHL*. Obtendremos así una teoría Δ escrita en lógica multivariada y parcial que representa —y rivaliza— con la original *XL*.

La novedad del planteamineto es que ahora, mediante un proceso de *retrotraducción*, vamos a definir una nueva lógica parcial *PXL* que será la que sustituya a la original *XL*. En esta fase debemos separar claramente las características parciales de las multivariadas. El resultado final es una lógica segura, matemáticamente mejor presentada y definida, una alternativa clara a *XL*.

COMENTARIO 5. *Mediante este procedimiento en (Huertas y Manzano, 1999) puede comprobarse cómo la modalidad de primer orden y la nueva parcialidad se encuentran de manera natural en una lógica modal y parcial. En este caso no se buscaba una reducción a la clásica, sino una nueva presentación de la lógica modal de primer orden, un* espejo *que nos devolviera una imagen mejorada. Caso de querer reducirla ahora, nos bastaría con usar los resultados de reducción de la lógica parcial.*

3. LÓGICAS GENERALES

3.1. *Introducción*

La teoría de las lógicas generales (Meseguer, 1989) es un estudio axiomático de los diferentes elementos de una lógica y sus interrelaciones; en particular, su relación de consecuencia sintáctica —integrada en el denominado *sistema implicativo*—, su cálculo y su aparato semántico —este último se estudia bajo la noción de *institución*.

La idea más interesante, extraída de la teoría de categorías, es que las propiedades más fecundas de un sistema lógico o de una estructura matemática no residen en ellos mismos, sino en su interrelación con sistemas o estructuras similares. Por ejemplo, se estudia mejor la teoría de grupos partiendo del hecho de que ellos y sus homomorfismos constituyen una *categoría*. Este punto de vista se adopta también en la teoría de las lógicas generales, en donde se considera que la consecuencia sintáctica, el cálculo deductivo y las denominadas *instituciones* son estructuras semejantes a los grupos del ejemplo y se establecen entre ellas correspondencias (*mapas*) capaces de preservar su estructura. Así que el punto fuerte de estas lógicas generales reside en definir adecuadamente las correspondencias entre ellas de tal manera que los *mapas* entre sistemas implicativos preserven la relación de consecuencia sintáctica, los establecidos entre los cálculos dejen fijo lo anterior y las propias pruebas y los que se definen entre instituciones conserven la relación semántica de satisfacibilidad.

En consonancia con esta *filosofía*, una lógica marco $\mathfrak{F}$ será aquella capaz de representar a muchas otras, y ello se concreta en el establecimiento de mapas de las lógicas en estudio a $\mathfrak{F}$.

3.2. *Definición: lógicas generales*

A continuación definimos brevemente sus componentes básicos.

3.2.1. Sintaxis

Para presentar un lenguaje se proporciona su signatura Σ y las reglas de formación de fórmulas. Para cada lógica hay una categoría *Sign* de sus posibles signaturas y un functor *sen* que a cada signatura Σ le asigna el conjunto de sus fórmulas, $sen(\Sigma)$.

3.2.2. Sistemas implicativos

Dada una signatura Σ en *Sign*, la derivabilidad de una sentencia $\alpha \in sen(\Sigma)$ a partir de un conjunto $\Gamma \subseteq sen(\Sigma)$ es una relación binaria que normalmente indicamos así: $\Gamma \vdash \alpha$, y que se establece siempre que a partir de los axiomas de Γ se sigan usando las reglas de la lógica en estudio —se dice que Γ *implica* α.

Formalmente se define un sistema implicativo como un triplo ordenado:

$$\mathcal{E} = \langle Sign, sen, \vdash \rangle$$

La relación $\vdash$ cumple reflexividad, monotonía y transitividad (corte) además de la propiedad de $\vdash$ —traducibilidad; esto es:

Si $\Gamma_{\Sigma}\ \alpha$ entonces para cada $H : \Sigma \rightarrow \Sigma'$ en *Sign* se cumple

$$Sen(H)(\Gamma) \vdash_{\Sigma'} sen(H)(\alpha)$$

Es decir, la derivabilidad se conserva aunque cambie el lenguaje. (A esto se le suele llamar principio de sustitución.)

3.2.3. Instituciones

Una institución es una cuádrupla

$$I = \langle Sign, sen, Mod, \Vdash \rangle$$

tal que *Sign* es una categoría cuyos objetos llamamos signaturas,

$$sen : Sign \rightarrow Set$$

es un functor que asocia a cada signatura un conjunto de sentencias,

$$Mod : Sign \rightarrow Set^{op}$$

es un functor que asocia a cada signatura una clase de modelos, y $\Vdash$ es una función que asocia a cada signatura Σ una relación binaria

$$\Vdash_{\Sigma} \subseteq Mod(\Sigma) \times sen(\Sigma)$$

que llamamos satisfacción, tal que para todo $\mathcal{M}' \in Mod(\Sigma')$, y para cada $\varnothing : \Sigma \rightarrow \Sigma'$, $\varphi \in sen(\Sigma)$, la siguiente propiedad, que se denomina *condición de satisfacción*, se cumple:

$$Mod(\phi)(\mathcal{M}') \Vdash_{\Sigma} \varphi \text{ syss } \mathcal{M}' \Vdash_{\Sigma'} sen(\phi)(\varphi)$$

La condición de satisfacción exige que para cada traducción sintáctica entre signaturas, ϕ, un modelo de la segunda signatura satisface la traducción de una fórmula si y sólo si la traducción del modelo satisface la fórmula original. El functor *Mod* es contravariante, funciona hacia atrás.

3.2.4. Consecuencia

Sea Γ un conjunto de Σ-sentencias. Entonces, se define $Mod(\Sigma, \Gamma)$ como el subconjunto de $Mod(\Sigma)$ determinado por todos los modelos $\mathcal{M} \in Mod(\Sigma)$ tales que satisfacen todas las sentencias de Γ, esto es, $\mathcal{M} \Vdash_{\Sigma} \varphi$ para cada $\varphi \in \Gamma$. Podemos ahora introducir una relación entre conjuntos de sentencias y sentencias, que denotamos $\Vdash$, del siguiente modo:

$$\Gamma \vDash_{\Sigma} \varphi \text{ syss } \mathcal{M} \Vdash_{\Sigma} \varphi \text{ para cada } \mathcal{M} \in Mod(\Sigma, \Gamma)$$

3.2.5. Mapa de instituciones

Sean las instituciones $\mathcal{I} = \langle Sign, sen, Mod, \Vdash \rangle$ e $\mathcal{I}' = \langle Sign', sen', Mod', \Vdash' \rangle$. Entonces, un mapa de instituciones $(\phi, \alpha, \beta) : \mathcal{I} \rightarrow \mathcal{I}'$ consta de una transformación natural $\alpha : sen \Rightarrow sen' \circ \phi$, un functor α-sensible $\phi : Th \rightarrow Th'$ entre las categorías de teorías de $\mathcal{I}$ e $\mathcal{I}'$, y una transformación natural $\beta : Mod' \circ \phi^{op} \Rightarrow Mod$, tal que para cada $\Sigma \in Sign$, $\Sigma \in sen(\Sigma)$ y $\mathcal{M}' \in Mod'(\phi(\Sigma,\varnothing))$ se cumple la siguiente propiedad:

$$\mathcal{M}' \Vdash_{sign'(\phi(\Sigma,\varnothing))} \alpha(\varphi) \text{ syss } \beta_{(\Sigma,\varnothing)}(\mathcal{M}') \Vdash_{\Sigma} \varphi.$$

3.3. *La lógica modal proposicional S4*

Dada la similitud de objetivos entre este planteamiento y el de mi libro (Manzano, 1996), parecía natural establecer vínculos. En Clavel y Manzano (1997) se realiza un primer esfuerzo para construir un puente entre sendos planteamientos que permita la comunicación de resultados ya establecidos en ambas orillas. Lo que hacemos en este artículo es centrarnos en una de las lógicas que habían sido satisfactoriamente traducidas a la multivariada, el sistema S4 de Lewis.

En primer lugar, definimos el aparato semántico de la lógica modal *S4* como una institución

$$\langle Sign_{PM}, Mod_{PM}, sen_{PM}, \Vdash \rangle$$

A tal efecto se define la categoría de signaturas modales proposicionales, que denotamos $Sign_{PM}$ y que tiene signaturas modales proposicionales como objetos y morfismos de signaturas modales proposicionales como morfismos; la composición y la identidad en $Sign_{PM}$ se expresan como composición de funciones y la función identidad, respectivamente.

¿Cuáles serán los modelos propios de esa lógica?

Un Σ-modelo de *S4* consta de: 1) un conjunto no vacío *W*, llamado el universo de estados o mundos; 2) una relación binaria *R*, $R \subseteq W \times W$, que es reflexiva y transitiva; y 3) una función $V : \Sigma \rightarrow \mathcal{P}(W)$, que asigna un conjunto de estados a cada proposición atómica y tal que $V(\bot) = \varnothing$.

Se requieren también los functores

$$Mod_{PM} : Sign_{PM} \rightarrow Set^{op}$$

y

$$sen_{PM} : Sign_{PM} \rightarrow Set$$

Para la introducción del concepto de verdad en un modelo de Kripke seguimos el procedimiento habitual, lo que nos lleva a establecer, por inducción sobre φ, que para todo morfismo $\phi : \Sigma \rightarrow \Sigma'$, y Σ'-modelo $\mathcal{M}'$,

$$Mod_{PM}(\phi)(\mathcal{M}') \Vdash_{\Sigma} \varphi \text{ syss } \mathcal{M}' \Vdash_{\Sigma'} sen_{PM}(\phi)(\varphi)$$

Por lo tanto se ha presentado la lógica *S4* como una institución.

A continuación se define un *mapa* de instituciones entre la lógica proposicional modal *S4* y la lógica multivariada. La $\alpha(\varphi, u) = \forall u TRANS(\varphi, u)$ —para la función *TRANS* se usa (Manzano, 1996)— define una transformación natural. Se precisa asimismo un functor Δ que envía:

— cada teoría modal proposicional (Σ, Γ) a una teoría multivariada de primer orden integrada por el conjunto $MODO(S4)_{\Sigma}$ y las traducciones de las fórmulas de la teoría (Σ, Γ);

— cada morfismo de teorías modales proposicionales a un morfismo de teorías multivariadas de primer orden.

Si añadimos a lo anterior la conversión de estructuras, $CONV_2$, el resultado es un *mapa* de *instituciones*.

COMENTARIO 6. *Vimos que hay tres niveles de* correspondencia *entre lógicas: 1) teorema de representación, 2) teorema principal, y 3) equivalencia de cálculos. El lema demostrado (en Clavel y Manzano, 1997) para* mapas de instituciones *establece el nivel de correspondencia entre lógicas del teorema de representación. Es de esperar que estos resultados puedan mejorarse en investigaciones futuras*[16].

16. En este mismo contexto se propone la *lógica de reescritura* como marco general y la herramienta informática *MAUDE* para realizar los cómputos; por aquí nos encaminamos hacia el «punto de encuentro».

4. SISTEMAS DEDUCTIVOS ETIQUETADOS

4.1. *Introducción*

El objetivo que Gabbay se propone es mostrar que la lógica clásica puede usarse como sistema universal, al menos desde el punto de vista de la deducción automática: existe un algoritmo que traduce casi cualquier lógica a la clásica.

Pese al inmovilismo que podría sugerir esta declaración de principios, lo que él hace es construir un marco lógico más flexible y próximo a las áreas de aplicación, investigando primero las necesidades que dichas aplicaciones suscitan.

Considera que la única forma de responder de forma convincente a la cuestión sobre la universalidad de la lógica clásica es investigar cuidadosamente cómo actúa y cuáles pudieran ser los procedimientos de traducción.

Su idea más revolucionaria es que la diferencia entre sistemas reside en el metalenguaje: la mayor parte de las lógicas están basadas en el *modus ponens*, las reglas de cuantificadores son las mismas; así que se distinguen en ciertos aspectos de su teoría de la prueba o de su semántica que se establecen desde fuera. Si se encontrase una lógica en la que las peculiaridades metalingüísticas se cifrasen en el lenguaje objeto —para ser recogidas en un cálculo— podríamos razonablemente esperar que ésta fuese un *marco general*. El que aparentemente mejor resuelve los problemas planteados es el de los sistemas deductivos etiquetados *LDS*.

Para apoyar su tesis se propone hacer dos cosas:

1. Mostrar que casi cualquier sistema lógico puede presentarse de forma natural como *LDS*. De hecho, algunas lógicas poseen requisitos funcionales que sólo se explicitan —y reciben mejor tratamiento— al ser traducidas.
2. *LDS* se traduce a lógica clásica de forma natural.

4.2. *¿Qué es un sistema lógico?*

Puesto que Gabbay concede especial importancia al sistema deductivo, una lógica no viene determinada simplemente por la relación $\vdash$ sino que se precisa entender también su peculiar procedimiento deductivo; así podríamos definir una lógica como un par: $\langle \vdash, S_{\vdash} \rangle$.

El paso siguiente se da al observar que el razonamiento humano no es estático, que agrega continuamente nuevos argumentos y que tampoco es monotónico. Ello le lleva a elaborar su primera idea de

un sistema deductivo etiquetado, concebido como marco en el que desarrollar lógicas. Se separa claramente del planteamiento tradicional cuando permite que los argumentos se sumen; esto se puede hacer en los *LDS*.

4.2.1. Sistemas lógicos como LDS

Una lógica se ha definido como un par $\langle \vdash, S_{\vdash} \rangle$ donde $\vdash$ es una relación de consecuencia estructurada —pudiera no ser monotónica, pero cumple alguna versión restringida de identidad y de corte— sobre un lenguaje *L* y $S_{\vdash}$ es un *LDS*.

Un sistema *LDS* es una tripla ordenada (*L*, *A*, *M*) donde *L* es un lenguaje lógico, *A* es un álgebra de etiquetas y *M* es la disciplina de etiquetado de fórmulas de la lógica. Contiene dos tipos de reglas: las de deducción y las de distribución y propagación de etiquetas.

Se definen los conceptos de *unidad declarativa*, *base de datos* y *etiqueta*. Las *etiquetas* atómicas son elementos de *A*, $t \in A$. Otras se forman recursivamente a partir de *A* usando los functores $f_1, ..., f_m$. Las *fórmulas* son las de *L*. Una *unidad declarativa* es un par de la forma $t : B$, donde *t* es una etiqueta y *B* es una fórmula; las etiquetas aportan información extra sobre la fórmula que no está codificada en ella. De hecho, las etiquetas funcionan como un sistema de control de las manipulaciones simbólicas a las que sometemos a las fórmulas en un proceso deductivo; guardan memoria de lo acontecido en el mismo. Las fórmulas etiquetadas aparecen en bases de datos que, a su vez, pueden estar etiquetadas, lo que permite que puedan intervenir en el proceso deductivo. Las *bases de datos*: son unidades declarativas o de la forma $\langle a, M, \mathrm{f} \rangle$, donde *M* es un diagrama finito; $\mathrm{a} \in M$ es una etiqueta singularizada y f asocia a cada etiqueta *t* de *M* o bien una base de datos o un conjunto finito de fórmulas. (Esta cláusula es recursiva. Una base de datos simple se obtiene siempre que la función f asocie a cada etiqueta un conjunto finito de fórmulas.) Las bases de datos pueden representar varias cosas: por ejemplo, un supuesto transitorio utilizado en una inferencia del cálculo de deducción natural para lógica clásica, o los mundos de la lógica modal o los estados de una máquina en lógica lineal, etc. El uso reglado de las etiquetas y lo que en ellas hayamos cifrado distingue a una lógica de otra.

Además, hay varias restricciones sobre el uso de *LDS*. En especial:

1. Sólo se usan reglas de inferencia clásicas: *modus ponens* y una versión del *teorema de la deducción*.

2. La distribución de etiquetas obedece reglas fijadas para toda lógica, pero aunque el formato sea fijo, permite variaciones de deta-

lle que hacen que incluso la regla de *modus ponens* se restrinja para ciertas lógicas.

3. Las variaciones entre lógicas que se producen en el metalenguaje aquí se expresan en el lenguaje objeto, en el mecanismo de etiquetado.

4.2.2. *Demostraciones en LDS*

En *LDS* se sustituye el concepto de prueba tradicional

$$B_1, \ldots, B_n \vdash C$$

por el de prueba de fórmulas con etiquetas

$$t_1 : B_1, \ldots, t_n : Bn \vdash s : C$$

Mientras que en la lógica tradicional las reglas de deducción se definen para fórmulas, en *LDS* se toman en consideración tanto ellas como sus etiquetas. Tenemos reglas para manipular las etiquetas, de forma que es ahí, en el álgebra de las etiquetas, en donde residen las características del metalenguaje responsables de las diferencias entre lógicas.

Ahora caracterizamos a un sistema lógico como a un par $\langle \vdash, LDS\vdash \rangle$ donde es una relación de consecuencia entre Δ —una base de datos etiquetada— y unidades declarativas, $t : A$. $LDS_{\vdash}$ es un sistema algorítmico para calcular la relación anterior.

4.3. *Semántica*

No se puede concluir sin mencionar al menos la semántica. El punto de vista tradicional, tanto para la clásica como para la intuicionista o la modal, es introducir el concepto de clase de modelos y el de evaluación de las fórmulas. Es decir, tenemos una clase $\mathfrak{K}$ de modelos y definimos $\mathcal{M} \Vdash A$ para cada modelo $\mathcal{M} \in \mathfrak{K}$ (validez de A en $\mathfrak{K}$). Si no entramos en el detalle de cómo son los modelos ni de cómo se evalúan las fórmulas, lo único que podemos decir es que con cada $\mathcal{M}$ se asocia una función que a cada fórmula le asigna un valor en el conjunto $\{0, 1\}$. Decimos que una lógica $\vdash$ es completa respecto de una clase $\mathfrak{K}$ de modelos si se verifica lo siguiente:

$$A \vdash B \text{ syss para cada } \mathcal{M} \in \mathfrak{K} \ (\mathcal{M} \Vdash A \Rightarrow \mathcal{M} \Vdash B)$$

La semántica de Gabbay, por el contrario, se incluye en los propios mecanismos de *LDS* y pasa a formar parte de la sintaxis. Esto se puede hacer de forma simple en el caso de la lógica modal, donde las etiquetas denotan mundos posibles y las reglas reflejan la semántica de las fórmulas. Por ejemplo, en la lógica *S4* la propagación de etiquetas respetará reflexividad y transitividad; las reglas para fórmulas podrán ser clásicas.

¿Qué se necesita para establecer de manera meramente sintáctica las características de la semántica? ¿Qué reemplazará a los modelos, a las interpretaciones, al propio concepto de completud?

Su respuesta es que una buena traducción entre lógicas suple a la semántica y a esto le llama *semántica sintáctica*. Sostiene que podemos prescindir de la semántica entendida en los términos de la teoría de modelos. Según su criterio, puede ser sustituida por una correcta traducción. Dados dos sistemas L_1 y L_2 decimos que L_2 puede ser usado como semántica para L_1 si el último puede ser traducido correctamente a L_2. De esta forma lo que correspondería a un teorema de completud sería el teorema de adecuación de la traducción (nuestro teorema 2.3.1).

Ejemplo 7. *Los siguientes pueden considerarse interpretaciones semánticas en el sentido de Gabbay*

1. *La interpretación de Boolos y Solovay de la lógica modal G (incluye el axioma de Löb) donde $\Box A$ significa «A es demostrable en la aritmética».*
2. *La interpretación de la lógica modal en la clásica.*

Para justificar y motivar su semántica Gabbay presenta a la tradicional de esta forma.

COMENTARIO 8. *Finaliza Gabbay su presentación de los* LDS *comparándola con la lógica clásica: el camino entre ambas es de ida y vuelta.*

— *Todo sistema* LDS *se puede traducir a la lógica clásica bivariada. Una de las variedades lo será para el álgebra de etiquetas. La traducción de $t : A$ es A^* (t). Éste es exactamente el procedimiento que seguimos en la traducción de la lógica modal. De esta forma el apoyo de la tesis de la universalidad de la lógica clásica viene dado por la generalidad de* LDS *y su papel unificador; no se necesita, según Gabbay, una interpretación semántica de la traducción. La elección del álgebra de etiquetas para representar a la lógica fue aquí lo verdaderamente determinante.*

— Es interesante señalar que a su vez la lógica clásica puede presentarse como un LDS *de forma no trivial.*

5. CONCLUSIONES

Intentaré sintetizar mi punto de vista. Para poder comparar sistemas lógicos es preciso saber qué son; ello nos conduce a una disyuntiva que se concreta en dos programas de investigación, posiblemente convergentes:

1. Ver qué problemas intentaban resolver con la esperanza de que la «*logicidad*» se oculte allí[17]; Este proyecto se retroalimenta, ya que conlleva, además de la reflexión sobre el concepto de lógica, la investigación de aquellas áreas de estudio emergentes del análisis que estén necesitadas de lógicas que las comprendan y expresen de forma natural, relevante, explicitando mediante reglas de la lógica sus raíces empíricas, pero exigiendo al producto no sólo coherencia interna —consistencia— sino todo el rigor metodológico y la precisión matemática al que la lógica nos tiene acostumbrados. Por su propia naturaleza, este ambicioso proyecto se puede poner en marcha, pero no efectuar su balance ahora, por lo que en el presente artículo no entra en consideración.

2. Pasando por alto el hacer previamente la definición comprensiva del concepto, usamos como criterio demarcador el meramente extensional —obtenido de la colección de *Handbooks*—, y tomamos una definición conceptual provisional como herramienta de trabajo. Analizamos qué posibilidades tenemos para comparar lógicas, considerando de interés los siguientes criterios:

a) Propiedades matemáticas de la relación de consecuencia.
b) Equilibrio entre potencia expresiva y deductiva.
c) Recursos matemáticos precisos para su desarrollo.
d) Traducibilidad a un marco común.

Así traducidas la comparación será sencillamente entre teorías de una misma lógica. El propio concepto de traducción y de lógica marco tiene que ser esclarecido; hemos optado (sección 2) por una traducción que permita que haya una correspondencia entre todos los componentes de las lógicas —modelos y cálculo— y hemos elegido a la lógica multivariada como marco.

17. Bajando la escalera, junto al *Aleph* de Borges.

5.1. *Reducción a la lógica multivariada*

El planteamiento está inspirado en la prueba de Henkin de completud de la teoría de tipos. Convertimos los modelos de la lógica en estudio *XL* en modelos extendidos en cuyos universos aparecen explícitamente todos los constructos definibles con las expresiones de *XL*. Para axiomatizar esta nueva clase de estructuras usamos la lógica multivariada, de la misma forma que en la prueba de completud mencionada. El *espíritu* de la correspondencia que se establece es *semántico*.

El programa se desarrolla a tres niveles, o etapas; en el primero nos detenemos en la mera representación de la verdad en la lógica en estudio como verdad de su traducción en estructuras multivariadas, módulo una cierta teoría cuya definición constituye una parte importante del programa de reducción; en el segundo la equivalencia se extiende al concepto de consecuencia a partir de conjuntos cualesquiera de hipótesis; en el tercero se demuestra la equivalencia del cálculo de la lógica en estudio con la teoría multivariada que la simula. Pese a su inspiración semántica, ya desde el primer nivel —incluso si la lógica *XL* careciera de cálculo—, podemos usar el cálculo multivariado para demostrar las fórmulas válidas de la lógica en estudio. La razón es que a partir del *teorema de representación* se demuestra fácilmente el *teorema de enumerabilidad* para la lógica *XL*. Por lo tanto, sabemos que podremos diseñar un cálculo para *XL*, pero también sabemos que en *MSL* podemos simular ese cálculo e incluso usar un *demostrador* de teoremas de *MSL*.

Así que, aunque la reducción se hubiera detenido en este nivel, los resultados desde un punto de vista práctico serían impresionantes.

Cuando la correspondencia supera también los niveles segundo y tercero obtenemos como subproducto los *teoremas de compacidad, Löwenheim-Skolem y completud* para la lógica *XL*.

En suma: pretendemos una unificación del *lenguaje máquina* y realizar en él todos los cómputos que en otro caso se harían en una multiplicidad de sistemas, y arrastrar las metapropiedades de la lógica multivariada a los sistemas lógicos en estudio. Sin embargo, mantenemos los diversos lenguajes como *lenguaje de usuario* por ser más naturales y adaptarse mejor a sus respectivos ámbitos de estudio.

5.2. *Otros planteamientos unificadores*

Se han estudiado otros planteamientos unificadores; a saber, el de las lógicas generales (sección 3) y el de los sistemas deductivos etiquetados (sección 4).

1. El análisis llevado a término en el primer caso es *supraestructural*, lo que quiere decir que se sitúa a un nivel de abstracción muy elevado. No nos extraña: la lógica subyacente es aquí la *teoría de las categorías*, que ocupa un nivel de abstracción superior al de la *teoría de modelos* que, a su vez, olvida la naturaleza de los objetos de la *realidad matemática* inmediata —números naturales, reales, etc.— y estudia las relaciones entre ellos. Una conceptualización de este tipo es matemáticamente muy poderosa, pero necesariamente más alejada del significado de las fórmulas.

2. En el segundo caso la maquinaria que se pone en funcionamiento es fundamentalmente sintáctica y a ella revierte la semántica. Se trata igualmente de una herramienta muy poderosa y de aplicabilidad en sistemas extraordinariamente diversos.

No deja de ser curioso que tanto cuando se utilizan en la traducción criterios *semánticos*, como cuando son *supraestructurales o sintácticos*, el resultado obtenido sea el mismo.

¿No será que la «logicidad» equivale a la traducibilidad a lógica clásica?

REFERENCIAS BIBLIOGRÁFICAS

Abramsky, S., Gabbay, D. y Maibaum, T. (1992-2000), *Handbook of Logic in Computer Science*, vol. 1 a 6, OUR, Oxford.

Alwein, G. y Barwise, J. (1996), *Logical Reasoning with Diagrams*, Oxford University Press, New York.

Barwise, J. y Etchemendy, J. (1994), *Hyperproof*, CSLI, Stanford.

Clavel, M. y Manzano, M. (1997), «Mapas de Instituciones: de la Lógica Modal S4 a la lógica multivariada», en A. Estany *et al.* (eds.) (1997), *Actas del II Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia*, Servei de Publicaciones de la UAB, Barcelona.

Fischer, M. y Ladner, R. (1979), «Propositional modal logic of programs»: *Proc. 9th ACM Symp. Theory of Comput.*, 18, 194-211.

Gabbay, D. (1994), *What is a Logical System?*, Oxford University Press, Oxford.

Gabbay, D., Hogger, G. y Robinson, J. (1993-1998), *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 1 a 6, OUP, Oxford.

Gabbay, D. y Guenthner, F. (²2001), *Handbook of Philosophical Logic*, Kluwer Academic Publishers, Dordrecht.

Gerbrandy, J. *et al.* (eds.), *JFAK: Essays dedicated to Johan van Benthem on the occasion of his 50th birthday*, Amsterdam University Press, Amsterdam.

Harel, D. (2001), «Dynamic Logic», en D. Gabbay y F. Guenthner, *Handbook of Philosophical Logic*, Kluwer Academic Publishers, Dordrecht.

Henkin, L. (1950), «Completeness in the theory of types»: *JSL*, vol. 15, 81-91.
Henkin, L. (1953), «Banishing the rule of substitution for functional variables»: *JSL*, vol. 18, n.º 3, 201-208.
Hodges, W. (1984), *Building models by games*, Cambridge University Press, Cambridge.
Huertas, A. y Manzano, M. (1999), «A fashionable partial and heterogeneous mirror for modality», en J. Gerbrandy *et al.* (eds.), *JFAK: Essays dedicated to Johan van Benthem on the occasion of his 50th birthday*, Amsterdam University Press, Amsterdam.
Huertas, A. (1994), *Modal Logic and Non-Classical Logic*, tesis doctoral de la Universidad de Barcelona.
Manzano, M. (1996), *Extensions of first order logic*, Number 19 in Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, Cambridge.
Manzano, M. (1998), «The little mermaid», en C. Martínez *et al.* (eds.), *Truth in Perspective*, Ashgate, Aldershot.
Meseguer, J. (1989), «General logics», en H.-D. E. *et al.* (eds.), *Logic Colloquium'87*, North-Holland, 275-329.
Van Benthem, J. (1986), «Correspondence theory», en D. Gabbay, y F. Guenthner, *Handbook of Philosophical Logic*, Kluwer Academic Publishers, Dordrecht.
Van Benthem, J. (1996), *Exploring Logical Dynamics*, CSLL, Stanford.

PROBLEMAS FILOSÓFICOS DE LA LÓGICA NO MONOTÓNICA

Raymundo Morado

En este artículo ofreceremos un panorama de los problemas filosóficos asociados al desarrollo de las lógicas no monotónicas. El objetivo es que la lectora entienda por qué surgen estas lógicas y tenga una idea clara de los desafíos que plantean a las nociones clásicas de inferencia. Se presuponen los conocimientos elementales de lógica cuantificacional, modal y no-monotónica que pueden encontrarse en textos como Quesada (1995), Orayen (1995) y Carnota (1995).

Después de ilustrar la historicidad de la noción de inferencia lógica, se hace un recuento de los límites teóricos y prácticos del modelo deductivo, con una clasificación de una familia de inferencias relacionados con la no-monotonicidad. A continuación se presenta un modelo más generoso de inferencia, con sus ventajas y sus propias limitaciones, y se ofrecen ejemplos.

Finalmente se presenta una formalización de las nociones tratadas, un recuento de propiedades formales de las inferencias no-monotónicas, se mencionan líneas abiertas de investigación y se incluye una bibliografía dividida en categorías para facilitar la investigación personal.

I. MOTIVACIÓN

1. *Historicidad de la noción de inferencia lógica*[1]

En cierto sentido la lógica no cambia; en otro sentido lo hace continuamente. No cambia en el sentido de que el *modus tollens* y el silo-

1. Esta parte está basada en Morado (2000a).

gismo disyuntivo son tan ciertos hoy como en los días de Crisipo. Cambia en sus estilos, énfasis, descubrimientos y problemas. Los nuevos usos de los sistemas lógicos llevan de manera natural a construir nuevos sistemas[2].

Entre los griegos la lógica fue la cristalización de una noción de racionalidad argumentativa. La silogística peripatética y la lógica proposicional estoica respondían tanto a preocupaciones ontológicas sobre la naturaleza de las cosas, como a una necesidad inmediata de construir y destruir argumentos. El uso retórico hizo de la lógica tanto una teoría de la argumentación, como un arte de discurrir y dialogar. El discurso más importante en la antigüedad clásica era el discurso legal y político, los alegatos que podían salvar o perder propiedades y vidas. Por ejemplo, el ciudadano ateniense que deseaba defenderse o demandar en persona (no, como hoy día, por boca de terceros), podía pagar a un experto en técnicas argumentativas para estructurar mejor su discurso. Y un discurso estaba bien estructurado cuando era convincente. El término «convincente» contiene una reveladora ambigüedad: puede significar algo psicológicamente irresistible para una persona de inteligencia normal o algo lógicamente irresistible. Estos dos sentidos son a menudo amalgamados en la mente griega y, como sus herederos, en pensadores posteriores hasta finales del siglo XIX, en que sobreviene el gran rechazo al psicologismo en ciencias formales. Desde los griegos, consideramos que la persona racional debe poder defender bien sus argumentos y atacar bien los de sus oponentes, en la doble vertiente de los *Analíticos Primeros* y las *Refutaciones sofísticas* de Aristóteles.

En la Edad Media el discurso paradigmático deja de ser el discurso forense. Ante la incertidumbre de salvar y proteger socialmente propiedad y cuerpo, el medieval intenta salvar individualmente su alma. El discurso más importante es el discurso religioso, que contiene las instrucciones reveladas para la salvación. El deseo de esclarecer la palabra divina ayuda a impulsar la reflexión hermenéutica. Entre los grandes lógicos medievales es común un interés en estudios de semántica y filosofía del lenguaje con teorías como las de los *insolubilia*[3]. Por supuesto que muchos de los intereses griegos permanecen. Cada uno de estos enfoques puede convivir con los anteriores, pero

2. Para una historia general de la lógica se pueden consultar los trabajos de Bochenski, 1962; Dumitriu, 1977; los Kneale, 1986; y Robles, 1995, mencionados en la Bibliografía.

3. Sobre las aparentemente insolubles paradojas lógicas. Por ejemplo, con Alberto de Sajonia.

hay una variación en el énfasis y en la presencia o ausencia de algunos temas específicos. Ahora la persona racional debe poder manejar las *obligationes*[4] del discurso y su análisis. El hombre racional es lógico en el sentido de que maneja y analiza bien el lenguaje.

La época moderna es asombrosamente pobre en desarrollos lógicos. La reacción contra el aristotelismo escolástico lleva a pensadores como Bacon, Descartes y sus seguidores a menospreciar la utilidad del instrumental lógico para enfatizar los nuevos métodos inductivos y experimentales. Durante este hiato modernista, la persona racional era la persona que sabía experimentar con la realidad física. Fuera de algunos intentos de Leibniz, publicados mucho después, este período tiene poco que ofrecer desde el punto de vista lógico.

Regresamos a los estudios lógicos con el siglo XIX, que toma la matemática como paradigma de racionalidad lógica. Este proceso culminó en la primera mitad del siglo XX con los grandes avances metalógicos sobre teoría de la cuantificación y teoría de conjuntos. Ya no se considera que la disciplina filosófica más cercana a la lógica es la ontología, como en Parménides, ni la retórica, como en Aristóteles, ni la filosofía del lenguaje, como en Ockham, sino la filosofía de las matemáticas, como en Frege y Russell. La persona racional debe poder deducir conclusiones a partir de verdades seguras como si desarrollara un sistema axiomático euclideano de geometría. La racionalidad se identifica con las llamadas funciones recursivas, como en la famosa «Tesis de Church». Hay quienes, trabajando bajo este nuevo enfoque, llegan a concebir el fin principal de la lógica como «la comprensión precisa y adecuada de la noción de *prueba matemática*»[5]. Éste es el enfoque «clásico» (para distinguirlo del «tradicional» que cubre desde Aristóteles hasta los escolásticos) y acompaña a los presupuestos actuales sobre lógica y racionalidad.

Desde la segunda mitad del siglo XX vivimos un nuevo cambio de enfoque. Gracias al éxito en su aplicación a las matemáticas, hemos visto extenderse el ámbito de aplicación de la lógica con aplicaciones en los campos más diversos. Con los nuevos usos se descubren limitaciones que antes no eran aparentes o importantes. Esto ha obligado a que los fines de la lógica se vuelvan más realistas y más complejos

4. Es decir, las «obligaciones» lógicas que contraemos cuando aceptamos considerar un caso hipotético, por mor del argumento, aunque asuma un supuesto contrario a los hechos. Para una discusión de la naturaleza y características de las *obligationes*, pueden verse, entre otros, los textos de Bradwardine, 1981; Fland, 1976, 1980; Spade, 1981, y Swyneshed, 1979.

5. «... a precise and adequate understanding of the notion of *mathematical proof*» (Mendelson, 1987, 1).

para poder incluir consideraciones de campos muy diversos como la lingüística o la filosofía del derecho. En las matemáticas se hacen más rigurosas las discusiones sobre lo que constituye una prueba gracias a las lógicas intuicionistas y constructivistas[6]. Las lógicas epistémicas[7] y modales[8] proporcionan modelos para discutir cuestiones clásicas sobre el conocimiento y la posibilidad. La física cuántica ha motivado el desarrollo de lógicas especiales[9], igual que el derecho inspiró investigaciones sobre lógicas deónticas[10]. En lingüística se usan desarrollos especiales de lógica cuantificacional para el estudio de las anáforas y la cuantificación ramificada. Los sistemas expertos en medicina utilizan lógicas condicionales y en ingeniería se recurre a lógicas polivalentes[11] y lógicas difusas[12]. Hay reconstrucciones del desarrollo de la química basadas en lógicas del descubrimiento (abductivas) y de la justificación (inductivas). En historia es útil la consideración de lógicas contrafácticas[13], y en pedagogía aparecen nuevas técnicas educativas de razonamiento crítico y lógica para niños.

Pero es en la disciplina más notoria de nuestra época, la ciencia y tecnología de la computación, donde se ha gestado el cambio más radical de nuestra idea de la lógica. La importancia de la lógica en esta área es múltiple. En informática, el cálculo proposicional auxilia en el diseño de circuitos lógicos, y el cálculo lambda en la teoría de la programación funcional. Hay desarrollos especiales aun cuando se utilicen elementos desarrollados inicialmente con otros objetivos en mente, como la teoría lógica de las funciones recursivas, o la teoría matemática de la complejidad. Se utilizan lógicas no monotónicas y técnicas de circunscripción para el manejo de la hipótesis de mundo cerrado en el manejo de bancos de datos, el problema del marco (*frame*) en robótica y el de la herencia en programación por objetos[14]. Las lógicas lineales se han usado en la construcción de semánticas para programas[15], y las temporales y dinámicas para control de sistemas. La informática tiene incluso su propio estilo de «programación lógica»,

6. Dos clásicos sobre este tema son Brouwer y Heyting.
7. Véase el libro de Hintikka mencionado en la Bibliografía.
8. Buenas referencias para lógica modal son Lewis y Langford, así como Hughes y Cresswell. También puede verse Hughes, Chellas, Morado y Beuchot, y Orayen.
9. Véase Jauch.
10. Son lectura obligada los libros de G. H. von Wright.
11. Su creador fue Jan Lukasiewics, en los trabajos mencionados en la bibliografía.
12. Creadas por Lofti Zadeh en su artículo de 1965.
13. Son sobre todo conocidas por el trabajo de David Lewis.
14. La bibliografía para estas lógicas es muy vasta. Los textos seminales son los de Reiter, McCarthy, McDermott, Moore, Shoam, Kraus, Lehmann y Magidor.
15. Véase, por ejemplo, Girard y Troelstra.

donde encontramos un concepto de negación como falla, así como técnicas lógicas especiales como la resolución y la unificación.

A los tradicionales modelos retóricos, exegéticos y matemáticos añadimos el del cálculo mecánico y efectivo. Hoy en día, la noción de racionalidad involucra computar bien, calcular o procesar eficientemente la información disponible. Como dijimos antes, cada uno de estos enfoques puede convivir con los anteriores. En estos tiempos, interesados en la racionalidad computacional, la teoría de funciones recursivas y la complejidad algorítmica, vemos una recuperación de los viejos ideales de la retórica, la semántica filosófica y la filosofía de las matemáticas. Y a todo esto hay que añadir un desarrollo reciente[16] de la inferencia contextual no deductiva. La persona lógicamente racional ya no será sólo «la que argumenta bien», ni «quien habla y comprende bien», ni «aquella que domina el álgebra del pensamiento», sino aquella que «procesa bien la información, dado su contexto». Todo lo anterior sobrevive y se extiende; todos esos intereses continúan y se entrelazan en distintas proporciones.

2. *Limitaciones del modelo clásico*

Ser «lógico» exige saber cuándo es adecuado ofrecer razones, cómo construirlas y cómo evaluarlas. Pero hay muchos tipos de razones. A principios del siglo XXI viejos modelos perviven con mínima tolerancia hacia el error o la revisión. En el modelo clásico para ser lógica una persona debe: 1) inferir fuera de contexto, 2) tener recursos ilimitados y ser lógicamente omnisciente, como si razonáramos fuera del tiempo y el espacio, y 3) ser infalible y consistente, sin necesidad de revisión. Veamos estas exigencias con más detalle.

La primera exigencia es que se asume el contexto límite en el que no hay presupuestos, el contexto vacío. Por ello es atractivo usar como unidades de análisis proposiciones entendidas como fragmentos de lenguaje desligados de las circunstancias de emisión. No preguntamos ¿cuándo?, ¿en qué sentido?, ¿para quién? El valor de verdad de una proposición no cambia por matices o cambios en las circunstancias, ya que las circunstancias no se consideran. De aquí que parezca natural asumir el compromiso específico de la bivalencia, según el cual sólo hay dos valores de verdad normalmente llamados «verdad» y «falsedad». La lógica y sus inferencias ocurren fuera del tiempo y el espacio.

16. Sobre todo desde principios de los años ochenta. Veremos esto con más detalle en la tercera parte.

Esta primera exigencia de acontextualidad lleva a la segunda exigencia. Se supone que el agente lógico es ideal y carece de límites en sus recursos para procesar la información. Si creemos A y de ello se sigue lógicamente B, entonces debemos creer B también. A esto le llamamos «omnisciencia lógica», pues se creen todas las consecuencias lógicas. La persona racional, en este paradigma, no deja de reconocer, y ciertamente nunca contradice, las consecuencias lógicas de sus creencias. Es natural que en la lógica clásica las consecuencias lógicas estén «cerradas» bajo la relación de inferencia[17]. Por ejemplo, Tarski pide que las consecuencias lógicas de las consecuencias lógicas no añadan nada, es decir, Cn(Cn A) = Cn(A).

Aunque una inferencia real requiere tiempo y espacio, para muchos lógicos clásicos el hecho de que un problema tenga *en principio* solución es todo lo que necesitan para considerarlo resuelto. Sin embargo, un problema resuelto en principio puede requerir de hecho más tiempo y espacio que el disponible para los agentes racionales existentes. Ser lógico se vuelve una exigencia inalcanzable para ser racional.

Esto nos lleva a la tercera exigencia. Además de la falta de contexto y limitantes en los recursos, se busca modelar solamente inferencias infalibles. Esto tiene dos ventajas: por un lado no hay necesidad de revisar creencias pasadas y por otro hay una garantía de que no habrá contradicciones si se parte de premisas consistentes. Un sistema axiomático que requiriera retracciones o cayera en contradicciones sería considerado inadecuado. En los modelos clásicos de sistemas axiomáticos se busca empezar sin errores[18]. A partir de esto se busca proseguir sin error, de manera deductiva. Con este modelo se excluye de entrada el tratamiento del error y la revisión.

Esta exigente visión de la lógica es un hermoso ideal para alcanzar, pero un peligroso criterio para juzgar sobre la racionalidad de un agente. Necesitamos un nuevo enfoque que considere de manera rigurosa, y si es posible formal, la estructura de la racionalidad en situaciones concretas, limitadas y falibles.

En el prólogo al *Begriffschrift*, Frege explica que su sistema está orientado sólo al contenido declarativo científico. Como vimos en la primera sección, eso se tradujo en términos de una de las ciencias más abstractas y rigurosas, la matemática. Al no caer en esta estrecha noción de conocimiento, los «matices» (*Farben*) tampoco caían dentro del campo de aplicación de la lógica.

17. En Tarski, 1935, 344.
18. El que la realidad sea otra, como con la famosa ley V de Frege, es visto como un desafortunado accidente.

La lógica clásica es adecuada para mundos ideales con recursos infinitos y sin cambio. La racionalidad clásica habla sobre la idealidad del agente, pero incluso un agente ideal puede vivir en condiciones no ideales. Como mencionamos antes, no es nuestra culpa que el mundo real cambia y los agentes crean, descubren, intercambian nueva información. Para ser racional puede bastar actuar no perfectamente sino tan perfectamente como las condiciones externas lo permitan. Esto es un cambio de óptica importante. Responder a los constreñimientos de la situación no es un menoscabo de la racionalidad mientras no haya defectos internos al pensamiento. Lo ideal sólo es normativo cuando es posible.

Al recolectar información, corremos riesgos y enfrentamos limitaciones. La información recabada puede estar equivocada, ser contradictoria, o ser incompleta. Necesitamos sistemas de lógica en los que se reconozca esto y que nos provean de mecanismos lógicos para manejarlo. Creo que el requisito de ser lógico es sensato, pero la lógica debe dar cabida a la sensatez falible. Esta sensatez no es un permiso para dejar de ser rigurosos, sino el intento de ser rigurosos en el contexto de nuestras limitaciones.

La lógica, pues, debe reconocer que podemos necesitar partir de inicios imperfectos, erróneos, contradictorios o incompletos. Y esto no es todo. Después de obtener la información, nuestro trabajo apenas empieza. La parte más importante es el procesamiento de los datos. Ese procesamiento, fuera de situaciones ideales, a veces exige saltar a conclusiones por razones teóricas, como la indecidibilidad, y prácticas, como la complejidad.

Aceptar que para ser racional se debe ser lógico no nos condena a sistemas lógicos inaplicables fuera de unas pocas situaciones afortunadas. Se puede ser lógico sin ser infalible mientras mantengamos inferencias que gocen de plausibilidad y sensatez. Estas nociones son vagas, pero la vaguedad se atempera gracias a que existen casos extremos claros y paradigmáticos.

Esta pérdida de la infalibilidad, reemplazándola con una modesta sensatez, no significa renunciar al rigor. Podemos incluso desarrollar sistemas que permiten y facilitan hacer revisiones a nuestros cuerpos de creencias, como las lógicas no-monotónicas en que es fácil modelar procesos de retracción de opiniones. Podemos tener sistemas formales paraconsistentes que acepten inconsistencias sin por ello caer en la trivialización de concluir cualquier cosa[19].

19. Sobre los cálculos paraconsistentes, véase da Costa y Marconi.

Para ser lógicos no necesitamos ignorar el contexto en que razonamos ni pretender que nuestros recursos son infinitos. Hay lógicas tetravalentes para modelar la manera en que cambia la evidencia sobre algo (ninguna, a favor, en contra, de ambos tipos). Las lógicas dinámicas permiten un manejo diacrónico de la inferencia, mientras que las lineales toman en cuenta que nuestra memoria y recursos similares para razonar son limitados.

La existencia de estos nuevos sistemas anuncia un nuevo concepto de racionalidad. Podemos tener todo el rigor formal de los sistemas clásicos sin sus presupuestos idealizadores. Tenemos al alcance una racionalidad para agentes de carne y hueso. Esta labor será larga y complicada por la variedad y complejidad de la experiencia humana. Una vez fuera del regazo protector de los axiomas y la deducción, tenemos que vérnosla con errores y planear su revisión.

Se abre el horizonte de toda una gama de tipos de razonamientos no deductivos. Aquellos en que se concluye una explicación (abductivos), los de sentido común aceptables para una comunidad, los *default* que se apoyan en lo que típicamente ocurre, los inciertos con reglas o premisas falibles, los inductivos con que se generaliza a partir de algunos casos, los no-monotónicos en general cuyas conclusiones son retractables a la luz de información adicional, los simplemente plausibles por estar altamente apoyados por la evidencia, y los *prima facie* a falta de información en contra. Son todos argumentos con algún grado de probabilidad, pero pueden ser bloqueados y sus conclusiones retractadas si el contexto cambia.

En esta nueva visión de la inferencia, aunque la lógica ideal no se aplique a todo en nuestras vidas, podemos ser lógicos y, gracias a ello, encontrar la racionalidad a nuestro alcance.

3. *Inferencias no deductivas*

Cotidianamente usamos los términos generales «razonamiento», «regla», «condicional» y «conclusión» como si fueran sinónimos. Igualmente con los adjetivos «de sentido común», «retractable», *default* («derrotable» o «por omisión»), «plausible», «no-monotónico» e «incierto». Sería deseable distinguir todos estos términos, por ejemplo, mediante la siguiente tabla:

	Razonamiento o argumento	*Inferencia, regla o condicional*	*Premisas*	*Conclusión*
De sentido común	Aceptable para la comunidad	Funciona sin presupuestos	Consideradas obvias	En armonía con creencias comunes
Retractable	Alcanza conclusiones retractables	Puede ser bloqueado si el contexto cambia	Retractable si el contexto cambia	Retractable si el contexto cambia
Default	Usa reglas *default*	Funciona a menos que sea inconsistente con la evidencia	Se asume verdad si no es inconsistente con la evidencia	Apoyada por una regla de *default*
Plausible	Alcanza conclusiones plausibles	Es apoyado por el contexto o evidencia más allá de cierto umbral	Es apoyado por la evidencia o contexto disponibles	Apoyada por reglas y premisas plausibles
No-monotónico[20]	Usa reglas no-monotónicas	Arroja una conclusión no-monotónica	No hay	Retractable si la información aumenta
Incierto	Usa reglas o premisas inciertas	Tiene excepciones (no es deductivo)	Falible o no confiable	Apoyada por reglas o premisas inciertas

Esta tabla no refleja todas las variadas formas en que estos términos han sido usados. Por ejemplo, «retractabilidad» (*defeasibility*) y no-monotonicidad se han usado como sinónimos[21]. Aunque esta tabla no ofrece definiciones rigurosas, refleja algunas distinciones importantes y evita que términos como *razonamiento retractable* confunda toda la gama de razonamientos no-aditivos, partes de inferencia de sentido común, inferencias *prima facie* y razonamientos falibles en general.

20. Hablo de «monotonicidad» y sus derivados en vez de «monotonía» para no confundir esta noción con la cercana (pero no idéntica) noción de monotonía matemática. Ésa es la práctica común en inglés, dónde se originaron estos estudios, e incluso autores en español que prefieren hablar «monótonamente» de vez en cuando hablan «monotónicamente» (*e.g.*, Carnota, 1995, 163).
21. Véase, por ejemplo, Dunlop y Fetzer (1993, 40).

También es importante evitar definiciones como «El razonamiento no-monotónico es razonamiento a partir de premisas verdaderas a conclusiones plausibles»[22]. Por un lado, nada nos impide inferir no-monotónicamente a partir de afirmaciones falsas y, por otro lado, si las conclusiones plausibles no son reversibles por evidencia adicional, el razonamiento es monotónico.

El razonamiento no-monotónico es retractable por un cambio en el contexto en términos de aumento de información (es deseable extender este concepto a argumentos con conclusiones no retractables). La conversa no es cierta: la retractabilidad no implica no-monotonicidad porque puede ser causada por varios tipos de cambio en el contexto, sólo uno de los cuales es incrementarlo. Esto nos impide extender el término «no-monotónico» a toda forma de razonamiento no deductivo.

Si nos sentimos generosos podemos incluir el razonamiento de sentido común como un tipo de razonamiento plausible, aunque inseguro. Nuestra decisión depende de cuán independiente de los estándares de la comunidad es nuestra noción de plausibilidad o apoyo evidencial. Pero sin importar lo que entendamos como reglas plausibles, algunas de ellas preservan verdad y otras no. Así pues, pueden ser tanto monotónicas como no-monotónicas; ciertamente las tautologías son por lo menos plausibles. Similarmente, aunque la mayoría de las reglas por *default* sean no-monotónicas, algunas no son retractables. Son los casos límite cuando las premisas implican la conclusión independientemente de la «justificación», en el sentido vacuo en que una conjunción implica ambos conyuntos «por *default*».

Hay que notar que la mitad de los tipos de inferencia en la tabla han sido clasificados de acuerdo a criterios formales. La noción de inferencia por *default* habla sobre consistencia, la no-monotónica carece de la propiedad lógica de cerrazón bajo reforzamiento de las premisas o antecedentes, y las inferencias inciertas se definen en contraste con la deducción. Este carácter formal las pone a un nivel distinto al de las inferencias de sentido común, plausibles o retractables en general.

Finalmente, podríamos tratar de subsumir al razonamiento por default, entendido como usando casos paradigmáticos o entendido como empleando una regla que sólo se aplica si no es contradicha por otra evidencia, en el razonamiento no-monotónico (razonamiento cuya conclusión puede tener que ser retractada a causa de un aumento en la información). Éste a su vez podría ser subsumido en un razonamiento retractable en que las retractaciones pueden ser motivadas no sólo por adición de información sino también por cambio o pérdida.

22. «Non-monotonic reasoning is reasoning from true premises to likely conclusions» (Stein, 1990, 1).

Las nociones de razonamiento retractable y de sentido común tienen un área de intersección, teniendo esta última noción un fuerte componente social. La percepción de la comunidad es crucial para decidir lo que hemos de llamar sentido común, pero si restringimos esta noción a un nivel psicológico podemos hablar de inferencias que parecen plausibles a algunos agentes bajo condiciones específicas.

Afortunadamente, partes de la lógica deductiva pueden ser llamadas plausibles y algunas incluso de sentido común. Por ende ninguna de estas dos nociones puede ser incluida en la de razonamiento incierto, aunque el tipo de razonamiento incierto que nos interesa es por lo menos plausible[23].

4. *Un concepto generoso de inferencia*

Las formas interesantes de retractabilidad en razonamiento por *default* se deben a su contextualidad. Al cambiar el contexto la inferencia deja de ser razonable. Los cambios en contexto cambian el grado de «razonabilidad» de las inferencias por *default*. Lo que hace a tales inferencias retractables no es la incertidumbre de las premisas o de la conclusión, sino el que su apoyo dependa crucialmente de su contexto. Así pues, en vez de representar relaciones por default sin relación con la base de creencias, debemos representarlas como relativas a esa base.

Hay un sentido de corregibilidad en la inferencia en el cual el razonamiento sigue siendo correcto aun después de ser cancelado. La retractabilidad no es falta de deducibilidad, pues es posible inferir por *default* algo necesario. La implicación no-monotónica puede ser el caso incluso cuando falla la material (y por ende la estricta y la contrafáctica).

Un condicional no-monotónico no queda invalidado por la información extra sino simplemente desactivado. La retractabilidad de la inferencia por *default* significa que es bloqueada por información adicional igual que las aseveraciones condicionales. Es decir, la nueva información no implica que la regla haya sido indebidamente aplicada sino que ya no se puede aplicar.

Un punto a enfatizar es que la inferencia es aceptable aun después de que nos damos cuenta de que la relación inferencial puede suspenderse al añadir información. Con el mismo cuerpo de creencias, aceptamos que X lleva a Y, y rechazamos que (X & Z) lleve a Y. La inferencia se mantiene aunque sabemos que depende de nuestra ignorancia.

23. Algunos autores reservan el término «incertidumbre» para hablar sobre errores numéricos.

Esto no significa que el contexto contiene suficientes premisas implícitas para hacer a la inferencia cierta si fueran hechas explícitas. Tal enfoque asimilaría a las inferencias no-monotónicas con argumentos con premisas tácitas que hubieran hecho a la inferencia monotónica si tan sólo hubieran sido explícitas. Ésta es la interpretación del entimema como un silogismo incompleto, con una inferencia clásica agazapada en el silencio. Aun teniendo todos el trasfondo de supuestos, la inferencia puede no ser segura, sólo altamente aceptable. La nueva información ni siquiera necesita contradecir al trasfondo de creencias para bloquear la inferencia de la conclusión, ya que no había en primer lugar ningún conjunto de supuestos que forzaban la conclusión tras bambalinas. En una inferencia no-monotónica nuestras premisas pueden ser insuficientes para la conclusión incluso después de haber tomado en cuenta los supuestos implícitos. Aun así el contexto puede hacer a la inferencia razonable aunque la nueva evidencia puede bloquear la inferencia previa al ofrecer fuerte evidencia para una tesis rival. Pero nunca tuvimos, ni lógica ni psicológicamente, la premisa implícita de que tal evidencia no existiera como quieren ciertos autores.

El razonamiento retractable no es necesariamente una regla irracional, ni requiere conclusiones equivocadas o premisas inseguras. Es una inferencia que depende del contexto y puede ser bloqueada. Por otro lado no se sigue de que las reglas sean revisables (pues la mayoría de las reglas científicas lo son), ni de que tenga premisas tácitas. No se trata de que el trasfondo sea verdadero y callado, sino de que sea aceptado y dependa del contexto.

Desde hace muchos siglos, el ideal del desarrollo del conocimiento ha sido deductivo. Una vez asegurado, un teorema permanece para siempre. Con esta noción de inferencia no tenemos que preocuparnos sobre el error, que sólo ocurre cuando abandonamos la segura senda de la deducción. Pero los desarrollos recientes en Inteligencia Artificial han enfatizado que los procesos deductivos de pensamiento son relativamente sencillos de reproducir, mientras que las inferencias diarias o de sentido común son mucho más difíciles de reconstruir, si hay poco lugar para el error y la revisión. Desde los años setenta se han multiplicado las voces que lamentan la transferencia del ideal acumulativo de las ciencias matemáticas[24].

Por supuesto, la lógica clásica ofrece algunos atisbos sobre la revisión racional de creencias. Nos da guías para añadir información

24. Un ataque muy influyente contra la monotonicidad en los sistemas lógicos apareció en Minsky (1974, 125).

con su noción de consecuencia lógica, e incluso para retractar información con los principios de *reductio ad absurdum* y *modus tollendo tollens*. Desafortunadamente también se enfatiza un modelo axiomático-deductivo del cambio racional de creencias en que simplemente añadimos creencias cuando la información aumenta, nunca las substraemos. Esta «aditividad» característica de la lógica deductiva clásica hace difícil explicar por qué y cómo abandonar creencias racionalmente ante nueva evidencia. ¿Qué podría ser tal evidencia? ¿Por qué forzar una pérdida de creencias previas? ¿Cuál sería una manera sensata de manejar tales pérdidas? No hay consenso entre los lógicos sobre las respuestas a estas preguntas, pero una lógica del razonamiento retractable (*defeasible*) podría ayudarnos.

Se esperaría que una teoría de la racionalidad ayudara a determinar qué debemos creer sobre la base de un cuerpo de evidencia dado. Pero lo que se sigue racionalmente de un cuerpo de creencias puede bien no ser lo que se sigue deductivamente de él. La racionalidad debe tomar en cuenta constreñimientos de nuestra situación. Por ejemplo, si el universo físico tiene un número finito de posibles estados, habría un infinito de proposiciones tautológicas que podrían ser en principio inferidas, pero no de hecho. En cierto sentido deberíamos inferirlas todas, pero si un deber epistemológico implica un poder epistemológico, no es claro que tenemos ese deber en un sentido epistemológico. Tal vez no fuera racional siquiera intentarlo.

Ya que una completa certeza en los fundamentos no es el caso normal, a menudo necesitamos saltar de nuestro conocimiento incompleto a conclusiones que avancen nuestras indagaciones. Los errores son un hecho de la vida diaria, tanto para humanos como para sistemas artificiales, para agentes aislados o para redes de ellos. Igual que lo necesitan los humanos, las máquinas necesitan ser capaces de modificar sus interpretaciones a la luz de nuevos datos que la máquina produce o encuentra.

Por ejemplo, para lograr eficiencia en la recolección de datos, las máquinas deben «conjeturar» el siguiente contorno o fonema que aparecerá en su campo visual o auditivo. Ya que las máquinas necesitan algo similar a nuestras creencias indexicales, puede ser menos costoso revisar tales creencias que generar todo un nuevo conjunto cada vez que algo cambia internamente o en el entorno[25].

Así pues, un sujeto epistémico capaz de enfrentar retos mínimos en el mundo real, sea una computadora o un humano, necesita poder manejar descripciones incompletas y/o inconsistentes sobre qué esta-

25. Estos puntos son argüidos en Doyle y London (1980, 7).

do de cosas ocurre. Normalmente usamos reglas que, aunque llevan a conclusiones retractables, garantizan un mínimo de racionalidad. La posibilidad del error conlleva la necesidad de retractar creencias.

Los humanos tenemos la capacidad de continuar operando a pesar de conflictos de creencias. Continuamos haciendo inferencias y tomando decisiones racionales en las áreas en que tenemos confianza y tan sólo dejamos de sacar (algunas) conclusiones en las áreas, tan pequeñas como sea posible, en que reconocemos la presencia de conflicto.

Los procesos racionales en situaciones de falibilidad son estudiados bajo diversos nombres como razonamiento inseguro, inferencia inductiva o consecuencia retractable. Hay que distinguir los casos en que nuestros datos pierden validez con el tiempo de los casos en que la información falta, es vaga o es poco confiable. Una medición precisa y confiable puede ser una función del momento en que la hacemos y, en este sentido, ser retractable porque el período de tiempo en el que funcionaba ha pasado. No se ha vuelto vaga o mal hecha, sino inoperante.

El precio de nuestra falibilidad es la necesidad ocasional de o bien recuperarnos de tales conflictos (*revisión de creencias*) o bien ser capaces de procesar datos adecuadamente en presencia de conflictos que pueden llegar a contradicciones. Ya sea que ajustemos o toleremos, debemos hacerlo siguiendo un mínimo de racionalidad. Creo que es tarea de la lógica determinar esta racionalidad.

El razonamiento retractable no sólo merece un estudio de su estructura lógica como una parte importante de la labor cognitiva de los sujetos epistémicos normales, sino también por razones prácticas. Desde Aristóteles los lógicos han enfatizado la deducción a expensas de la inducción pero hasta hoy las técnicas lógicas han sido lentas e incompletas. Perder información o habilidades puede incluso ser conveniente en ciertas condiciones[26].

Al admitir la retractabilidad admitimos perder confiabilidad, pero esto es razonable en algunas áreas. Todavía usamos humanos en muchas industrias no porque sean más baratos o precisos que un robot, sino porque son más rápidos. El tiempo disponible para responder a un ataque nuclear puede ser demasiado pequeño para las elaboradas cadenas de inferencias usadas en nuestros actuales probadores de teoremas.

26. Lesgold, Glaser, Rubinson, Klopfer, Feltovich y Wang (1988) reportan que en ciertos contextos de aprendizaje (el andar de un bebé, la actuación de un radiologista) se pierde parte de una habilidad para desarrollar otra más tarde. Senger (1989, 88-89) hace la hipótesis de que algo similar puede ocurrir con el desarrollo del razonamiento legal en estudiantes de Derecho.

Así pues, tanto desde un punto de vista teórico como práctico, a menudo hay que negociar con el error.

5. *Algunos ejemplos*

Consideremos los siguientes ejemplos:

Tweety. Se le dice a usted que Tweety es un pájaro y usted concluye que Tweety vuela (Reiter, 1980, 68).

Café. Usted cree que si pone azúcar a su café tendrá buen sabor. Concluye que si pone azúcar y aceite para auto en su café, tendrá buen sabor (ter Meulen, 1986, 138).

Aerolínea. Se le dice a usted que Airline Canada vuela de Vancouver a Toronto, Boston y Los Ángeles. Cuando otra persona le pregunta si vuela a Toulouse, usted dice que no (Reiter, 1978, 301).

Nixon. Del hecho que Nixon es cuáquero usted infiere que es un pacifista. Del hecho que es un republicano usted infiere que no es un pacifista[27].

Robot. Después de dejar caer un bloque rojo, un robot asume que el color del bloque no ha cambiado[28].

Paracaídas. Un hombre cayó de un avión. Afortunadamente, llevaba puesto un paracaídas. Desafortunadamente, el paracaídas no se abrió. Afortunadamente, cayó del avión a baja altura sobre un enorme montón de heno. Desafortunadamente, había un tridente en el montón de heno. Afortunadamente, no cayó sobre el tridente. Desafortunadamente, tampoco cayó sobre el montón de heno...[29].

Estos ejemplos ilustran importantes problemas. El ejemplo de Tweety muestra que modos perfectamente normales y sensatos de razonar son falibles o retractables. Si descubriéramos que Tweety es un pingüino o que tiene un ala rota, ya no podríamos concluir que vuela. La importancia de este ejemplo de juguete no tiene nada que ver con la ornitología y tiene todo que ver con las intuiciones de que conclusiones razonables pueden ser falibles (el ejemplo del café), que normalmente tomamos información incompleta como si fuera completa (el ejemplo de la aerolínea), que de la misma información podemos inferir conclusiones en conflicto (el ejemplo de Nixon), y que nuestros ajustes de creencias se dan en el contexto de un mundo independiente y cambiante (el ejemplo del robot).

27. Reiter y Criscuolo (1981, 98). El sujeto del ejemplo fue llamado «John» ahí.
28. Un ejemplo más complejo de este tipo es analizado en McCarthy y Hayes (1969, 36-37).
29. Contado en Nute (1990, 351).

Estos problemas han recibido creciente atención desde finales de los años sesenta por sus repercusiones sobre ciencias de la computación, inteligencia artificial y lógica filosófica. A veces los ejemplos dejan ver sus orígenes en la teoría de banco de datos (el ejemplo de la aerolínea), cibernética (el del robot) o contrafácticos (el ejemplo del paracaídas). Pero no debemos dejarnos engañar por la novedad de la formulación; estas cuestiones han estado con nosotros desde los comienzos mismos de la lógica.

6. *Formalización*

La lógica deductiva clásica reconoce que en ocasiones llegamos a conclusiones inaceptables. La figura estoica del *modus tollendo tollens*[30] nos dice que la conjunción de las premisas (o el antecedente) sufre similarmente de falsedad. Algo paralelo hace la *reductio ad absurdum* con la falsedad necesaria. Pero en ambos casos lo que la lógica clásica nos da para resolver un conflicto es una negación, no una retracción. No corregimos el error, tan sólo lo ubicamos, identificando una combinación de creencias a evitar. En la tradición lógica, el método favorito de solución es la abstinencia[31].

Esta noción de la consecuencia lógica ha sido cuidadosamente descrita. Tarski caracteriza al conjunto de Consecuencias Lógicas de un conjunto Γ (simbolizado como «$Cn(\Gamma)$») como el conjunto de oraciones que podemos inferir de Γ[32]. Esto se entiende como una noción primitiva que podemos aplicar incluso a teorías que carecen de una definición rigurosa de inferencia. No hemos especificado la teoría de prueba correspondiente a la noción de inferencia ni hemos determinado el lenguaje en que tiene lugar. Por ejemplo, no sabemos si el lenguaje tiene conjunciones o si contamos con una regla de simplificación. Aun así, la noción de consecuencia lógica debe cumplir ciertas condiciones. Tarski las captura en la siguiente serie de axiomas[33]. Siendo S el conjunto de todas las oraciones,

30. Estoica por ser uno de los cinco «modos indemostrables» de Crisipo; véase Kneale y Kneale, 1986, 162-163.

31. Arnauld y Nicole, 1662, expresan esta actitud: «il y a une infinité d'esprits grossiers et stupides que l'on ne peut reformer en leur donnant l'intelligence de la verité, mais en les retenant dans les choses qui sont à leur portée, et les empeschant de iuger de ce qu'ils ne sont pas capables de connoitre».

32. Tarski (1930a). En sentido estricto, Tarski habla en partes del artículo sobre tipos de oraciones, pero ignora esta sutileza no esencial en este contexto, cosa que yo también haré. Posteriormente favoreció la definición de sistemas deductivos en términos del conjunto de todas las oraciones válidas (como noción primitiva) en vez de Cn. Véase Tarski, 1935, 344.

33. A estos axiomas generales podemos añadir otros especiales para la negación y

[Axioma 1] La cardinalidad de $S \leq \aleph_0$.
[Axioma 2] Si $\Gamma \subseteq S$, entonces $\Gamma \subseteq Cn(\Gamma) \subseteq S$.
[Axioma 3] Si $\Gamma \subseteq S$, entonces $Cn(Cn(\Gamma)) = Cn(\Gamma)$.
[Axioma 4] Si $\Gamma \subseteq S$, entonces $Cn(\Gamma) = \Sigma_{\Delta \subseteq \Gamma \text{ \& la cardinalidad de } \Delta < \aleph_0} Cn(\Delta)$.
[Axioma 5] Hay una oración $\Gamma \in S$ tal que $Cn(\Gamma) = S$.

El axioma 1 nos dice que el número de nuestras oraciones debe ser contable, un requisito impuesto por la primera diagonal de Cantor dada la restricción normal a concatenaciones finitas de símbolos.

El axioma 2 garantiza una forma plausible del principio de identidad $A \rightarrow A$. Llamaré a este axioma Inclusividad[34].

El axioma 3 nos dice que un conjunto de consecuencias es un punto fijo. Éste no es un supuesto trivial, pues las inferencias «inmediatas» (de un paso) o las «obvias» no lo respetan. Este axioma también implica que nuestra noción de consecuencia no está limitada por constreñimientos en nuestros recursos, como límites espaciales o temporales. Llamo a esto la Cerrazón Lógica del operador de consecuencia.

El axioma 5 dice que hay una oración «total» que implica todas las demás.

He dejado para el final el axioma 4, llamado usualmente Compacidad. Me concentraré en este axioma porque de él obtenemos inmediatamente el teorema Si $\Gamma \subseteq \Delta \subseteq S$, entonces $Cn(\Gamma) \subseteq Cn(\Delta)$. Tarski dice explicitamente que «la operación Cn en el dominio de los conjuntos de oraciones es monotónica»[35].

Es difícil argüir con el principio de monotonicidad $Cn(\Gamma) \subseteq Cn(\Gamma \cup \Delta)$ cuando $\Delta \subseteq Cn(\Gamma)$. Esto ha sido llamado Monotonicidad Cauta y con el principio de restrictividad $\Gamma \subseteq Cn(\Delta) \rightarrow Cn(\Gamma \cup \Delta) \subseteq Cn(\Delta)$ obtenemos que las consecuencias de Γ son ni más ni menos que las consecuencias de su conjunción con cualquiera de sus consecuencias.

el condicional material, pero éstos se aplican sólo cuando deseamos presuponer el cálculo de oraciones bivalente clásico, así es que no los consideraré.

34. Tomo el término Inclusividad (inclusiveness) de Wojcicki (1988). Wojcicki define una operación de consecuencia como aquella que satisface Inclusividad más la propiedad $\Gamma \subseteq Cn(\Delta) \rightarrow Cn(\Delta \cup \Gamma) \subseteq Cn(\Delta)$ que él llama Restrictividad (restrictiveness). La intuición parece ser que añadir consecuencias lógicas no incrementa el poder inferencial (véase el axioma 3). La noción tarskiana (y la monotonicidad) aparecen al añadir el axioma 4.

35. «The operation Cn in the domain of sets of sentences is monotonic», Tarski (1930b, 65), y menciona otras dos formas equivalentes: $\Sigma_{\Gamma \in \Re} Cn(\Gamma) \subseteq Cn(\Sigma_{\Gamma \in \Re} \Gamma)$ y $Cn(\Pi_{\Gamma \in \Re} \Gamma) \subseteq \Pi_{\Gamma \in \Re} Cn(\Gamma)$. Es decir, la unión de las consecuencias está en las consecuencias de la unión, y las consecuencias de la intersección están en la intersección de las consecuencias.

Combinaré ahora las nociones tarskianas de Inclusividad, Cerrazón y Compacidad con otros principios para obtener cuatro interesantes nociones de consecuencia y mostrar cómo extender cada una en la que sigue. Empiezo con un operador de Consecuencia Mínima que obedece dos principios:

[Inclusividad] $\Gamma \subseteq Cn(\Gamma)$
[Restrictividad] $\Gamma \subseteq Cn(\Delta) \rightarrow Cn(\Gamma \cup \Delta) \subseteq Cn(\Delta)$.

La Inclusividad nos dice que del cuerpo de evidencia Γ se sigue al menos Γ mismo. Por supuesto, hay sentidos en que esto no es el caso. Después de aceptar un cuerpo de evidencia con fuertes conflictos (tal vez internos, tal vez incluso lógicos), podemos con buenas razones inclinarnos no a inferir la evidencia misma sino a rechazar parte de ella. Éste es un uso legítimo de la palabra «consecuencia» que excluimos con este principio.

El principio de Restrictividad nos dice que añadir consecuencias no incrementa el poder inferencial de una teoría. De nuevo, hay un sentido en el que añadir como premisas ciertas consecuencias hace posible obtener más consecuencias (tal vez en un sentido temporal o psicológico). Incluso la lógica y las matemáticas se apoyan en descubrimientos previos. Tal noción de consecuencia, aunque también razonable, no será capturada por nuestros operadores de Consecuencia Mínima.

Ahora, un operador de consecuencia Cumulativo añade el principio

[Cumulatividad] $\Gamma \subseteq Cn(\Delta) \rightarrow Cn(\Delta) \subseteq Cn(\Gamma \cup \Delta)$

Es decir, tiene Inclusividad más $\Gamma \subseteq Cn(\Delta) \rightarrow Cn(\Delta) = Cn(\Gamma \cup \Delta)$ que es Restrictividad y Cumulatividad juntas. Nótese que sin el antecedente $\Gamma \subseteq Cn(\Delta)$, el consecuente $Cn(\Delta) \subseteq Cn(\Gamma \cup \Delta)$ sería Monotonicidad. Estamos en efecto diciendo que la consecuencia cumulativa se comporta monotónicamente con respecto a la adición de información *vieja*. Esto es una «cauta monotonicidad». Ya que $Cn(\Gamma \cup \Delta)$ no es sino $Cn(\Delta)$ misma, no hay un verdadero aumento de información de que preocuparse.

Una Consecuencia Monotónica respeta los siguientes tres principios:

[Inclusividad] $\Gamma \subseteq Cn(\Gamma)$
[Cerrazón] $Cn(Cn(\Gamma)) \subseteq Cn(\Gamma)$
[Monotonicidad] $\Gamma \subseteq \Delta \rightarrow Cn(\Gamma) \subseteq Cn(\Delta)$

Finalmente, un Operador de Consecuencia Tarskiano simplemente añade a los otros dos principios de la Consecuencia Mínima lo siguiente:

[Compacidad] $\alpha \in Cn(\Gamma)$ ssi para un $\Delta \subseteq \Gamma$ finito, $\alpha \in Cn(\Delta)$

En breve:

Mínima	= Inclusividad	+ Restrictividad		
Cumulativa	= Inclusividad	+ Restrictividad	+ Cumulatividad	
Monotónica	= Inclusividad	+ Cerrazón	+ Monotonicidad	
Tarskiana	= Inclusividad	+ Restrictividad	+ Compacidad	

Substituí Cerrazón con Restrictividad para mayor claridad. Podemos hacer esto porque, como se mostrará en un momento, dada Inclusividad, la Restrictividad implica Cerrazón y la Restrictividad puede, a su vez, ser derivada de las propiedades de los operadores de consecuencia monotónicos. Hay que hacer notar que el par de propiedades de la Consecuencia Mínima (Inclusividad y Restrictividad) no pueden ser reemplazadas con el par de Inclusividad y Cerrazón. El segundo par es implicado por el primero porque, como acabamos de mencionar, la Inclusividad y la Restrictividad implican Cerrazón, pero no al revés. Piénsese en Cn como una función sobre conjuntos de reales que arroja al conjunto original más su promedio, su promedio más 3 y su promedio menos 3. Este operador respeta Inclusividad porque su resultado incluye al conjunto original, y respeta Cerrazón porque el promedio nunca cambia, pero no respeta Restrictividad: sean $\Delta = \{ 1, 9 \}$ y $\Gamma = \{ 8 \}$. Ya que $Cn(\Delta) = \{ 1, 2, 5, 8, 9 \}$, tenemos $\Gamma \subseteq Cn(\Delta)$, pero $Cn(\Gamma \cup \Delta) = \{ 1, 3, 6, 8, 9 \}$ no es un subconjunto de $Cn(\Delta)$.

La relación entre los cuatro operadores de consecuencia es la siguiente[36]:

Mínima ▶ Cumulativa ▶ Monotónica ▶ Tarskiana.

Prueba:

I. Mínima ▶ Cumulativa. Por construcción.

II. Cumulativa ▶ Monotónica. La Inclusividad es compartida, así es que sólo necesitamos mostrar que la Restrictividad y la Cumulati-

36. «A ▶ B» significa que cualquier operador de consecuencia de tipo B es un operador de consecuencia de tipo A. La prueba de que ninguno de estos ▶s puede ser contrapuesto se deja al legendario lector interesado.

vidad pueden ser derivadas de la Inclusividad, Cerrazón y Monotonicidad. Asumimos $\Gamma \subseteq Cn(\Delta)$. Para mostrar Restrictividad necesitamos $Cn(\Gamma \cup \Delta) \subseteq Cn(\Delta)$. De nuestro supuesto $\Gamma \subseteq Cn(\Delta)$ y el hecho de que $\Delta \subseteq Cn(\Delta)$ (por Inclusividad), obtenemos $(\Gamma \cup \Delta) \subseteq Cn(\Delta)$. A partir de esto Monotonicidad arroja $Cn(\Gamma \cup \Delta) \subseteq Cn(Cn(\Delta))$. Pero por Cerrazón esto se simplifica como $Cn(\Gamma \cup \Delta) \subseteq Cn(\Delta)$. Para mostrar Cumulatividad necesitamos $Cn(\Delta) \subseteq Cn(\Gamma \cup \Delta)$, que se obtiene fácilmente de Monotonicidad, ya que $\Delta \subseteq (\Gamma \cup \Delta)$.

III. Monotónica ▶ Tarskiana. Nuevamente, la Inclusividad es compartida, así es que sólo tenemos que mostrar que Inclusividad, Restrictividad y Compacidad implican Cerrazón y Monotonicidad. De hecho, la Inclusividad y la Restrictividad son suficientes para la Cerrazón, mientras que la Compacidad es suficiente para la Monotonicidad. Para mostrar Cerrazón, empezamos con $Cn(\Gamma) \subseteq Cn(\Gamma)$, lo que Restrictividad transforma en $Cn(Cn(\Gamma) \cup \Gamma) \subseteq Cn(\Gamma)$. Pero sabemos por Inclusividad que $(Cn(\Gamma) \cup \Gamma) = Cn(\Gamma)$. Por lo tanto, $Cn(Cn(\Gamma)) \subseteq Cn(\Gamma)$, que es Cerrazón. Ahora, para Monotonicidad, asumimos $\Gamma \subseteq \Delta$ y mostramos que $Cn(\Gamma) \subseteq Cn(\Delta)$. Así pues, asumimos también que $\alpha \in Cn(\Gamma)$. Por Compacidad, $\alpha \in Cn(\Gamma)$ para algún subconjunto finito Ξ de Γ. Pero $\Xi \subseteq \Delta$ por nuestro primer supuesto. Por ende, para algún subconjunto finito Ξ de Δ, $\alpha \in Cn(\Delta)$. Por Compacidad[37], $\alpha \in Cn(\Delta)$. QED.

8. *Propiedades formales de las inferencias no-monotónicas*

Una amplia gama de fenómenos pueden ser relacionados con no-monotonicidad. Esto ha producido diferentes enfoques, aparentemente inconexos. Afortunadamente, casi todos estos fenómenos comparten importantes propiedades formales y el estudio de un tipo puede iluminar el estudio de otros.

Decimos que una función f es monotónica bajo un orden O ssi O(x,y) implica O(fx, y). En el caso de razonamiento monotónico, O es la relación «implica» y f es cualquier función que añada contenido semántico. En otras palabras, añadir información a las premisas debiera preservar la conclusión. *E.g.*, $(\Phi \rightarrow \Psi)$ sólo si $[(\Phi \wedge X) \rightarrow (\Psi)]$. (En términos de uniones de conjuntos de piezas de información: Si $\Gamma \leq \Delta$ entonces $\Gamma \cup \Xi \leq \Delta$.)

37. Como veremos en un momento, este último y crucial paso ha usado una noción fuerte de Compacidad.

Tarski habla de la operación monotónica Cn que arroja un conjunto de consecuencias a partir de un conjunto de oraciones (Tarski, 1930b, 65), y la lógica clásica, lo mismo que las intuicionistas y modales, son llamadas monotónicas porque la adición de información no afecta a la validez de la inferencia deductiva (véase Gabbay, 1985, 439).

Esto es formalmente análogo a la regla de Debilitación que aparece en lógica combinatorial como el combinador K. En lógica algebraica corresponde al principio de límite inferior $a \circ b \leq b$. En algunos sistemas de deducción natural (o cálculos de secuencias tipo Gentzen) Debilitación aparece como

$$\frac{\Gamma, \Delta \vDash \Phi}{\Gamma, \Psi, \Delta \vDash \Phi}$$

Por otro lado, una relación de orden O no-monotónica es aquella que viola la regla de que si $\Phi \leq \Psi$ bajo O, entonces $g(\Phi) \leq g(\Psi)$, para cualquier función g que incremente contenido semántico. Por ejemplo, si substituimos $g(\alpha)$ por $(\alpha \wedge X)$, tenemos que $\Phi \leq \Psi$ no implica $(\Phi \wedge X) \leq (\Psi \wedge X)$. Incrementar nuestro conocimiento de Φ a $\Phi \wedge X$ puede impedirnos inferir Ψ.

La analogía entre monotonicidad y Debilitación no es perfecta, pero ha llevado a aplicar el término no-monotonicidad al rechazo de Debilitación. Otra caracterización es decir que el razonamiento monotónico ocurre cuando las inferencias se preservan bajo aumento de premisas: las reglas de inferencia son siempre de la forma *Φ es un teorema si $\Psi_1, ..., \Psi_n$ son teoremas*. En contraste, el razonamiento no-monotónico ocurre cuando las inferencias no se preservan bajo aumento de premisas: las reglas de inferencia pueden ser de la forma *Φ es un teorema si $\Psi_1, ..., \Psi_n$ no son teoremas*. Se dice que las reglas de inferencia hacen a los axiomas *permisivos* o *restrictivos*[38].

Hay que notar que hay otras operaciones no-monotónicas que no se relacionan directamente con incremento de información. Por ejemplo, podemos definir una noción de *consecuencia mínima* como[39]

$$\Gamma \vDash_m \Phi \text{ ssi } \forall M \, ([M \vDash \Gamma \wedge \forall N \, (N \vDash \Gamma \rightarrow N \not\subset M)] \rightarrow M \vDash \Phi)$$

38. Minsky, 1974, 125, se quejaba: «En cualquier sistema de lógica, todos los axiomas son necesariamente "permisivos"; todos contribuyen a permitir que se obtengan nuevas inferencias (In any logistic system, all the axioms are necessarily "permissive" —they all help to permit new inferences to be drawn)». Véase también Moore, 1985.

39. Papalaskari y Weinstein, 1990, 3. Véase también Papalaskari, 1988.

Es decir, cualquier modelo mínimo de Γ modela Φ. Si definimos modelos como conjuntos de oraciones atómicas Φ, Ψ, ..., y estipulamos que $\Gamma \vDash \neg\Phi$ ssi $\Phi \notin \Gamma$[40], es claro que $\varnothing \vDash_m \neg\Phi$, pero $\varnothing \cup \{ \Phi \} \vDash_m \neg \Phi$. Esta consecuencia mínima es no-monotónica en el sentido de que superconjuntos de los modelos pierden consecuencias. Pero, por nuestra estipulación sobre $\neg$, los superconjuntos no se producen por un mero incremento de conocimiento, pues determinan explícitamente todo lo que es el caso e implícitamente todo lo que no. Agrandar estos modelos representa un incremento de información «positiva» tanto como un decremento de la información «negativa». En contraste, la noción de no-monotonicidad que buscamos es no-monotonicidad bajo incremento total de información, tanto positiva como negativa.

Debemos distinguir dos formas de debilitar una relación de consecuencia no-monotónica $\approx$: mediante el reforzamiento de las premisas o aligerando la conclusión. Cada variante añade material a la derecha o a la izquierda de $\approx$. La Debilitación a la Izquierda[41] $\Phi \approx \Psi$, $\Gamma \vdash \Phi \rightarrow \Gamma \approx \Psi$ implicaría monotonicidad (basta reemplazar Γ con $\Phi \wedge X$). Por otro lado, la Debilitación a la Derecha es aceptable: $\Phi \approx \Psi$, $\Psi \vdash \Gamma \rightarrow \Phi \approx \Gamma$.

Es decir, la inferencia no-monotónica está cerrada bajo deducibilidad clásica. La justificación intuitiva de Debilitación a la Derecha es que si algo es implicado infaliblemente, también es implicado faliblemente. Podemos pensar en la inferencia no-monotónica como una relación encima de la deducibilidad clásica, por lo que llamaremos a esta propiedad Supraclasicalidad[42]: $\Phi \vdash \Psi \rightarrow \Phi \approx \Psi$.

Supraclasicalidad nos dice que estamos extendiendo el sistema de inferencia clásica. La relación de inferencia $\approx$ es derivable a partir de un condicional material válido, pero no tenemos aún un modo de iterar esto.

La Transitividad (o Silogismo Hipotético) $\Phi \approx \Psi$, $\Psi \approx X \rightarrow \Phi \approx X \rightarrow \Phi \approx X$ debe ser evitada: supóngase $\Phi \approx \Psi$. Ya que $\Phi \wedge X \approx \Phi$ (por Supraclasicalidad), $\Phi \wedge X \approx \Psi$ por Transitividad[43], y hemos vuelto a Debilitación.

Los condicionales no-monotónicos han sido interpretados como diciendo que la mayoría de los individuos de cierto tipo tienen cierta

40. Como hacen, por ejemplo, Chang y Keisler, 1973, 4-5.

41. Llamada Restricción del Antecedente (Antecedent Restriction) en Adams, 1975, 17.

42. Otra relación que exhibe Supraclasicalidad es el condicional material.

43. Aparentemente el principio de Transitividad está tan internalizado que Glymour y Thomason, 1984, 96, lo perdieron de vista completamente y culparon en su lugar al *Modus Ponens* por esta inferencia.

característica. Una ventaja de esta lectura es que el cuantificador «la mayoría de» no permite Contraposición interna. Decir que todos los Φs son Ψs es equivalente a decir que todos los no-Ψs son no-Φs por contraposición: $\Phi \approx\!\!| \Psi \rightarrow \neg\Psi \approx\!\!| \neg\Phi$. Pero la información de que la mayoría de los Φs son Ψs no es lo mismo que decir que la mayoría de los no-Ψs son no-Φs. Morreau sostiene que una situación similar ocurre con los *defaults* y que esta falla de contraposición ayuda a resolver la llamada paradoja de la inducción, donde un tigre amarillo confirma que los cuervos son negros. Cuando la ley inductiva es una regla por *default*, la paradoja de la inducción no surge porque la contraposición falla. La «paradoja» permanece para cuantificaciones universales *reales* (Morreau, 1988, 339).

Dudo que la falla de contraposición ayude a resolver la «paradoja» de la inducción, pues me parece que este famoso problema no es sobre condicionales. Simplemente subraya la importancia de elementos heurísticos que a veces se pasan por alto al discutir en abstracto estrategias de confirmación. Efectivamente, es una mala idea no circunscribir nuestros tests a los cuervos, que son menos numerosos que las cosas que no son negras, pero no es lógicamente incorrecto. De todos modos, Morreau tiene razón en señalar que la contraposición debe fallar para reglas por *default*, cualquiera que nuestro análisis final de ellas sea.

La regla de ∧-Introducción $\Phi \mathrel{|\!\approx} \Psi, \Phi \mathrel{|\!\approx} X \rightarrow \Phi \mathrel{|\!\approx} (\Psi \wedge X)$ parece inocente a pesar de ejemplos como el de Nixon. Sabemos que los republicanos por *default* no son pacifistas, y que los cuáqueros son por *default* pacifistas. Pero no nos han dicho qué es por *default* un republicano cuáquero.

La importante regla de Corte debe restringirse. Corte Completo: $(\Omega \wedge \Psi) \mathrel{|\!\approx} X \rightarrow (\Omega \wedge \Phi) \mathrel{|\!\approx} X$ implicaría Transitividad. Asúmase que $\Phi \mathrel{|\!\approx} \Psi$ y $\Psi \mathrel{|\!\approx} X$. Si Ψ es idempotente con $\Psi \wedge \Psi$ (o si ‹∧› se entiende como unión de conjuntos de premisas), tenemos $(\Psi \wedge \Psi) \mathrel{|\!\approx} X$ y $\Phi \mathrel{|\!\approx} \Psi$ a partir de lo cual Corte Completo arroja $(\Psi \wedge \Phi) \mathrel{|\!\approx} X$. Por Corte Completo nuevamente obtenemos $(\Phi \wedge \Phi) \mathrel{|\!\approx} X$, lo cual colapsa en $\Phi \mathrel{|\!\approx} X$, y tendríamos Transitividad después de todo.

Veamos esto con calma. Otra forma de Corte, Corte Cauto $\Phi \mathrel{|\!\approx} \Psi, (\Phi \wedge \Psi) \mathrel{|\!\approx} X \rightarrow \Phi \mathrel{|\!\approx} X$, dice que si X es implicado por una proposición Φ en conjunción con un Φ implicado de todas maneras, entonces X es implicado por Φ solo. Creo que esto es aceptable porque una proposición está *implícitamente* conjuntada con todo aquello que implica[44].

44. Nótese la similitud con Transitividad. Adams, 1975, 22, rechaza que Transitividad sea probabilísticamente aceptable, pero acepta esta forma como universalmente

Corte Completo dice algo completamente diferente: si algo implicado por Φ implica X con la ayuda de Ω, entonces Φ debe ser capaz de usar Ω para implicar lo mismo. Tal principio no funciona para la inferencia no-monotónica. Lo que algo implicado por Φ puede implicar a solas podría no ser implicado más cuando el resto del poder de Φ es invocado. ¿Quiere esto decir que nunca es *kosher* reforzar premisas? No. Monotonicidad Cauta $\Phi \mid\!\approx \Psi, \Phi \mid\!\approx X \rightarrow (\Phi \wedge \Psi) \mid\!\approx X$ nos dice que añadir información implícita es monotónico.

Compacidad es un caso interesante. La entenderé como compacidad de una consecuencia lógica («Si Φ es una consecuencia de Γ, es también una consecuencia de algún subconjunto finito de Γ»), lo cual en el caso clásico (de primer orden) coincide con compacidad de la satisfacibilidad («Si cada subconjunto finito de Γ es satisfacible, lo es Γ»)[45]. Hay también una noción de Compacidad Débil que simplemente dice que nuestra noción de prueba es finitista:

$\Gamma \mid\!\approx \Phi \leftrightarrow \exists \Delta (\Delta \subseteq \Gamma \wedge$ Cardinalidad de $\Delta < \aleph_0 \wedge \Delta \mid\!\approx \Phi)$. Podemos aceptar, de momento, que todas las implicaciones no-monotónicas surgen de un conjunto finito de premisas. Lo que no podemos aceptar es que eso es todo lo que la implicación representa, como en Compacidad Fuerte:

$\Gamma \mid\!\approx \Phi \leftrightarrow \exists \Delta (\Delta \subseteq \Gamma \wedge$ Cardinalidad de $\Delta < \aleph_0 \wedge \Delta \mid\!\approx \Phi)$. Es fácil ver que la Compacidad Fuerte produciría monotonicidad, ya que cualquier subconjunto finito de Γ es también un subconjunto finito de todos los superconjuntos de Γ[46].

Otras propiedades estructurales clásicas no tocan directamente el asunto de la no-monotonicidad, por lo que es natural darlas por aceptadas. Por ejemplo, Inclusividad ($\Phi \mid\!\approx \Phi$), que ya conocemos; o, por meras consideraciones de teoría de conjuntos, tenemos contracción ($\Phi, A, \Psi, A, \Omega, \mid\!\approx X \rightarrow \Phi, A, \Psi, \Omega \mid\!\approx X$) y Permutación ($\Gamma, \Phi, \Psi, \Omega \mid\!\approx X \rightarrow \Gamma, \Psi, \Phi, \Omega \mid\!\approx X$).

Para resumir, podemos esperar que una relación binaria no-monotónica $\mid\!\approx$ carezca de las siguientes propiedades:

Monotonicidad	$\Phi \mid\!\approx \Psi \rightarrow (\Phi \wedge X) \mid\!\approx (\Psi \wedge X)$
Debilitación	$\Phi \mid\!\approx \Psi \rightarrow (\Phi \wedge X) \mid\!\approx \Psi$
Debilitación Izquierda	$\Phi \mid\!\approx \Psi, \Gamma \vdash \Phi \rightarrow \Gamma \mid\!\approx \Psi$

probabilísticamente aceptable y la llama Silogismo Hipotético Restringido (*Restricted Hypothetical Syllogism*).

45. Éstas son dos versiones comunes del Teorema de Compacidad de Gödel-Malcev; véase Barwise, 1977, 10.

46. Un hecho mencionado en Tarski, 1930b, 64.

Transitividad	$\Phi \mid\!\approx \Psi, \Psi \mid\!\approx X \rightarrow \Phi \mid\!\approx X$
Corte Completo	$\Phi \mid\!\approx \Psi, (\Omega \wedge \Psi) \mid\!\approx X \rightarrow (\Omega \wedge \Phi) \mid\!\approx X$
Compacidad Fuerte	$\Gamma \mid\!\approx \Phi \leftrightarrow \exists\Delta\ (\Delta \subseteq \Gamma \wedge$ Cardinalidad de $\Delta < \aleph_0 \wedge \Delta \mid\!\approx \Phi)$

pero que tenga al menos las siguientes propiedades:

Monotonicidad Cauta	$\Phi \mid\!\approx \Psi, \Phi \mid\!\approx X \rightarrow (\Phi \wedge \Psi) \mid\!\approx X$
Supraclasicalidad	$\Phi \vdash \Psi \rightarrow \Phi \mid\!\approx \Psi$
Debilitación Derecha	$\Phi \mid\!\approx \Psi, \Psi \vdash \Gamma \rightarrow \Phi \mid\!\approx \Gamma$
$\wedge$-Introducción	$\Phi \mid\!\approx \Psi, \Phi \mid\!\approx X \rightarrow \Phi \mid\!\approx (\Psi \wedge X)$
Corte Cauto	$\Phi \mid\!\approx \Psi, (\Phi \wedge \Psi) \mid\!\approx X \rightarrow \Phi \mid\!\approx X$
Compacidad Débil	$\Gamma \mid\!\approx \Phi \rightarrow \exists\Delta\ (\Delta \subseteq \Gamma \wedge$ Cardinalidad de $\Delta < \aleph_0 \wedge \Delta \mid\!\approx \Phi)$

9. *Problemas abiertos*

Quedan por tratar problemas filosóficos importantes sobre los formalismos no-monotónicos. Mencionaré algunos sobre los que se está trabajando actualmente: ¿en qué sentido podemos hablar de lógicas sin validez perfecta?, ¿cuáles son los compromisos filosóficos de estas nuevas nociones de consecuencia lógica?, ¿cuál es el estatus en esos formalismos de reglas como el *modus ponens*?, ¿se le rechaza verdaderamente? ¿Hasta qué punto es legítimo introducir consideraciones prácticas sobre las limitaciones de agentes finitos en la construcción de una teoría de la inferencia lógica? ¿Deben ser los sistemas no-monotónicos declarativos, formales, simbólicos? ¿Cuáles son las unidades adecuadas de análisis en estas lógicas: creencias, conocimientos, oraciones, aseveraciones? ¿Debe la conclusión de una inferencia no-monotónica aparecer en todo el rango de extensiones consideradas, o hasta qué punto se legitima la conclusión por aparecer al menos en una extensión razonable? ¿Cuáles son los requisitos para una formalización aceptable de la noción de contexto para los fines de una teoría general de la inferencia? ¿Cuál es el mentalenguaje adecuado para estos formalismos, y cómo integrar el uso de nuevas reglas y nuevos axiomas con el efecto de los cambios de contexto? ¿Cuál es la relación exacta entre inactivación de reglas y supuestos y su negación?

REFERENCIAS BIBLIOGRÁFICAS

I. *Algunos trabajos en español*

Beck, A. (1996), «Paradojas clásicas en el contexto de la Lógica no monótona»: *Agora*, vol. 15, n.º 2, 145-152.

Carnota, R. J. (1995), «Lógica e Inteligencia Artificial», en *Enciclopedia Iberoamericana de Filosofía*, vol. 7, *Lógica*, Trotta, Madrid, 143-183.

Casaban, E. (1989), «Lógica y razonamiento no monotónicos»: *Quaderns de Filosofía de la Ciencia*, vol. 15-16, 19-25.

Fariñas del Cerro, L. y Frías Delgado, A. (1995), «Razonamiento no monótono: Un breve panorama»: *Theoría, S. E.*, vol. X, n.º 23, 7-26.

Morado, R. (1992), «Models for belief revision», en E. Villanueva (ed.), *Rationality in Epistemology*, Ridgeview, Atascadero, 227-247.

Morado, R. (2000a), «Nuevos paradigmas de la inferencia racional», en C. Trueba (comp.), *Racionalidad: Lenguaje, Argumentación y Acción*, UAM-I/Plaza y Valdés, México, 89-99.

Morado, R. (2000b), «La justificación retórica de los principios lógicos», en H. Beristáin y M. Beuchot, *Filosofía, Retórica e Interpretación*, colección «Bitácora de Retórica», UNAM, 163-176.

Núñez, G. y Cortes, U., «Implicaciones del conocimiento de respaldo en el aprendizaje SBL y su relación con el razonamiento no monótono», 237-261.

II. *Clásicos de las lógicas no-monotónicas*

Gabbay, D. M. (1985). «Theoretical foundations for non-monotonic reasoning in expert systems», en R. A. Krzysztof (ed.), *Logics and Models of Concurrent Systems; Proceedings of the NATO Advanced Study Institute on Logics and Models of Concurrent Systems held at La Colle-sur-Loup*, France, 8-19 octubre 1984, NATO ASI series. Series F, *Computer and System Sciences*, vol. 13, Springer-Verlag, Berlin, 439-457.

Gärdenfors, P. (1988), *Knowledge in Flux*, MIT Press, Cambridge.

Gärdenfors, P. (1991), «Belief revision and nonmonotonic logic: Two sides of the same coin? Abstract», en JELIA (1991), 52-54.

Guha, R. V. (1990), «The representation of defaults in Cyc», en AAAI (1990), 608-614.

Guha, R. V. (1991), *Contexts: A formalization and some applications*, Ph. D. Thesis, Stanford.

Konolige, K. (1986), *A Deduction Model of Belief*, Morgan Kaufmann.

Konolige, K. (1987), *On the relation between default and autoepistemic logic*, CSLI Report n.º CSLI-87-105. También en IJCAI (1987), en *Artificial Intelligence*, vol. 35, issue 3, 343-382; (1988), y en Ginsberg (1987), 195-226.

Konolige, K. (1989), «On the relation between circumscription and autoepistemic logic», en IJCAI (1989), 1213-1218.

Konolige, K. (1989), «Errata to: On the relation between circumscription and autoepistemic logic»: *Artificial Intelligence*, vol. 41, 115.

Kraus, S., Lehmann, D. y Magidor, M. (1990), «Non-monotonic Reasoning, Preferential Models and Cumulative Logics»: *Artificial Intelligence* 44, 167-207.

McCarthy, John (1977), «Epistemological problems of artificial intelligence», en IJCAI, 1038-1044.

McCarthy, J. (1980), «Circumscription: A form of non-monotonic reasoning»: *Artificial Intelligence* 13, 27-39. (Reimpreso en Ginsberg [1987], 145-151.)

McCarthy, J. (1980b), «Addendum: circumscription and other non-monotonic formalisms»: *Artificial Intelligence*, vol. 13(1-2), 171-172. (Reimpreso en Ginsberg [1987], 151-152.)

McCarthy, J. (1984), «Applications of circumscription to formalizing common sense knowledge», en NMRIW (1984), 295-324. Versión revisada en *Artificial Intelligence* 28(1), 89-116, 1986, y reimpresa en Ginsberg (1987), 153-166.

McCarthy, J. (1993), «Notes on formalizing context», en IJCAI (1993), 555-560.

McCarthy, John y Hayes, P. (1969), «Some philosophical problems from the standpoint of artificial intelligence», en B. Meltzer y Michie (eds.), *Machine Intelligence*, vol. 4, Edinburgh University Press, Edinburgh, American Elsevier, New York, 463-502. (Reimpreso en B. L. Webber y N. J. Nilsson [eds.], *Readings in Artificial Intelligence*, Tioga, Palo Alto, California, 1981, 431-450, y en McCarthy, 1990, *Formalization of Common Sense: Papers by John McCarthy*, editado por V. Lifschitz, Norwood, N. J., Ablex. Un resumen apareció en el ejemplar de febrero de 1969 (n.º 14) de la ACM SIGART Newsletter. (Reimpreso en Ginsberg [1987], 26-52.)

Mcdermott, D. V. y Doyle, J. (1978), «Non-monotonic logic I», MIT Technical Report Memo 486. También *Artificial Intelligence*, vol. 13(1-2), 41-72, 1980. Un resumen amplio apareció en el Fourth Workshop on Automated Deduction, Austin, Texas, 1979. (Reimpreso en Ginsberg [1987], 111-126.)

McDermott, D. V. (1982), «Non monotonic logic II: Non monotonic modal theories»: *Journal of the Association for Computing Machinery*, vol. 29, n.º 1, 33-57.

Minsky, M. L. (1974), *A framework for representing knowledge. Artificial Intelligence Memo 306*, MIT AI Lab. (Reimpreso sin el apéndice en *The Psychology of Human Vision*, P. H. Winston [ed.], McGraw Hill, New York, 1975, 211-277.) Otros fragmentos fueron reimpresos como «Frame-system theory» en R. C. Schank y B. L. Webber (eds.), *Proceedings of the 1975 TINLAP (Theoretical Issues in Natural Language Processing) Conference*, Cambridge, Massachusetts, 1975, reimpreso en P. N. Johnson-Laird y P. C. Wason (eds.), *Thinking: Readings in Cognitive Science*, Cambridge University Press. Hay extractos distintos en J. Haugeland (ed.), *Mind Design: Philosophy, Psychology, Artificial Intelligence*, Montgomery VT, Bradford Books, 95-128.

Moore, R. C. (1985), «Semantical considerations on nonmonotonic logic»: *Artificial Intelligence*, vol. 25, n.º 1, 75-94. (Reimpreso en Ginsberg

[1987], 127-136.) Una versión más breve apareció en IJCAI (1983), 272-279. Véase también SRI AI Center Technical Note 284 (1983).
Reiter, R. (1980), «A logic for default reasoning»: *Artificial Intelligence*, vol. 13, n.º 1-2, 81-132. (Reimpreso en Ginsberg [1987], 68-93.)
Shoam, Y. (1986), «Chronological ignorance: An experiment in nonmonotonic temporal reasoning»: *AAAI*, 389-393, y en Ginsberg (1987), 396-409.
Shoam, Y. (1987), «A semantical approach to nonmonotonic logic»: *IJCAI* (1987), 388-392, y Ginsberg (1987), 227-250. Resumen extenso en LICS (1987), 275-279.
Shoam, Y. (1987), «Nonmonotonic logics: Meaning and utility», en IJCAI (1987), 388-393.
Shoam, Y. (1988), *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*, MIT Press, Cambridge, Mass. Básicamente, su tesis de Yale.

III. *Textos introductorios*

Antoniou, G. (1997), *Nonmonotonic Reasoning*, MIT Press, Cambridge, Mass.
Besnard, P. (1989), *An Introduction to Default Logic*, Springer Verlag, Alemania, esp., 1-5 y 31-36.
Besnard, P., Brewka, G., Froidevaux, Ch., Gregoire, E. y Siegel, P. (1991), «Nonmonotonicity»: *Journal of Applied Non-Classical Logics*, vol. 1, n.º 2, 267-310.
Brewka, G. (1989), «Nonmonotonic logics –a brief overview»: *AI Communications*, vol. 2, n.º 2, 88-97.
Brewka, G. (1991a), *Nonmonotonic Reasoning: Logical Foundations of Commonsense*, Cambridge Tracts in Theoretical Computer Science 12, Cambridge University Press, New York.
Brewka, G., Dix, J. y Konolige, K. (1991), «A tutorial on nonmonotonic reasoning», en *Nonmonotonic and Inductive Logic: Second International Workshop*, Reinhardsbrunn Castle, Germany, 2-6 diciembre 1991.
Brewka, G., Jantke, K. P. y Schmitt, P. H. (eds.) (1993), *Lecture notes in computer science, 659, Lecture notes in artificial intelligence*, Springer-Verlag, Berlin/New York, 1-88.
Brewka, G., Dix, J. y Konolige, K. (1997), *Nonmonotonic Reasoning: An Overview*, Center for the Study of Language and Information. Lecture Notes, Number 73, Stanford, California.
Donini, F. M., Lenzerini, M., Nardi, D., Pirri, F. y Schaerf, M. (1990), «Nonmonotonic Reasoning»: *The Artificial Intelligence Review*, vol. 4, n.º 3, 163-210.
Etherington, D. W. (1988), *Reasoning With Incomplete Information*, Morgan Kaufmann Publishers, London, especialmente 1-73 y 157-163.
Geffner, H. A. (1992), *Default Reasoning: Causal And Conditional Theories* MIT Press, especialmente 1-22.
Ginsberg, M. L. (1987), *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Los Altos, California.

Lukascewicz, W. (1985), *Non-Monotonic Reasoning: Formalization of Common-sense Reasoning*, Ellis Horwood, Chichester, esp., 77-88.
Marek, V. W. (1993), *Truszczynski, Miroslaw, Nonmonotonic Logic: Context-Dependent Reasoning*, Springer-Verlag, Berlin.
McCarthy, J. (1995), *CS323: Nonmonotonic Reasoning*, disponible por FTP en www.formal.stanford.edu/jmc/, sobre todo los capítulos 6, 8 y 10.
Minker, J. (1993), «An overview of nonmonotonic reasoning and logic programming», por FTP en ftp.cs.umd.edu:pub/prism/lpnmr.overview.jlp.ps. También en *The Journal of Logic Programming*, vol. 17, n.º 2/3/4, 95, y como Technical Report UMIACS-TR-91-112, CS-TR-2736, U. of Maryland at College Park, MD, 1991.
Perlis, D. (1987), «Reasoning, nonmonotonic», en Shapiro, S. C. (ed.), *Encyclopedia of Artificial Intelligence*, Wiley, New York, 848-853.
Sombe, L. (Phillipe Besnard, Marie-Odile Cordier, Didier Dubois, Luis Fariñas del Cerro, Christine Froidevaux, Yves Moinard, Henri Prade, Camilla Schwind y Pierre Siegel) (1990), *Reasoning Under Incomplete Information in Artificial Intelligence: A Comparison of Formalisms Using a Single Example. Special Issue of the International Journal of Intelligent Systems*, vol. 5, n.º 4, 323-472.

IV. *Bibliografía de apoyo o citada en este capítulo*

Adams, E. W. (1975), *The Logic of Conditionals*, Reidel.
Arnauld, A., Nicole, P. (1662), *La Logique ou L'Art de Penser*, Ch. Savreux, Paris.
Barwise, J. (1977), «An introduction to first-order logic», en J. Barwise (ed.), *Handbook of Mathematical Logic*, 5-46. Cito de la séptima impresión, 1991.
Bochenski, I. M. (²1962), *Formale Logik*, K. Alber, Freiburg. Traducción de Millán Bravo Lozano, *Historia de la Lógica Formal*, Gredos, Madrid, 1966.
Bradwardine (1981), «*Insolubilia* and Bradwardine's *Theory of Signification*», editado por Paul Vincent Spade, *Medioevo* VII, 115-134.
Carnota, R. J. (1995), «Lógica e Inteligencia Artificial», en *Enciclopedia Iberoamericana de Filosofía*, vol. 7: *Lógica*, Trotta, Madrid, 143-183.
Chang, C. C. y Keisler, H. J. (1973), *Model Theory*, North-Holland Publishing Company.
Chellas, B. F. (1980), *Modal Logic*, Cambridge University Press, London.
da Costa, Newton C. A. (1974), «On the theory of inconsistent formal systems»: *Notre Dame Journal of Formal Logic*, vol. XV, n.º 4, octubre, 497-510.
da Costa, Newton C. A. (1982), «The Philosophical Import of Paraconsistent Logic»: *The Journal of Non-Classical Logic*, vol. I, n.º 1, 1-19.
Doyle, J., London, P. (1980), «A selected descriptor-indexed bibliography to the literature on belief revision»: *SIGART Newsletter*, n.º 71, 7-23.
Dumitriu, A. (1977), *History of Logic*, vols. I, II. Abacus Press, Tunbridge Wells, Kent. Traducción revisada, actualizada y aumentada de la segunda edición de *Istoria Logicii*, Editura Didactic, Bucarest, 1975. Traducido por Duiliu Zamfirescu, Dinu Giurc Neanu y Doina Doneaud.

Dunlop, Ch. E. M. y Fetzer, J. H. (1993), *Glossary of Cognitive Science*, Paragon House, New York.
Fland, R. (1976), «Robert Fland's *Consequentiae*: An Edition», editado por Paul Vincent Spade, *Mediaeval Studies* 38, 54-84.
Fland, R. (1980), «Robert Fland's *Obligationes*: An Edition», editado por Paul Vincent Spade, *Mediaeval Studies*, vol. XLII, 41-60.
Girard, J.-Y. (1987), «Linear Logic»: *Theoretical Computer Science* 50, 1-102.
Glymour, C. y Thomason, R. H. (1984), «Default reasoning and the logic of theory perturbation», en *Non-Monotonic Reasoning Workshop*. Organizado por la American Association for Artificial Intelligence, 17-19 octubre 1984. Mohonk Mountain House, New Paltz, New York, 93-102.
Hintikka, J. (1989), *The Logic of Epistemology and the Epistemology of Logic*, Kluwer, Dordrecht.
Hughes, G. E. (1984), *A companion to Modal Logics*, Methuen, London.
Hughes, G. E. y Cresswell, M. J. (1973), *Introducción a la lógica modal*, Tecnos, Madrid. Original (1972) *An Introduction to Modal Logic*, Methuen, London. (Primera edición 1968.)
Jauch, J. M. (1968), *Foundations of Quantum Mechanics*, Addison Wesley.
Kneale, W. C. y Kneale, M. (1986), *The development of logic*. Reimpresión con correcciones, Clarendon, Oxford. Traducida la 3.ª edición por J. Muguerza como *El Desarrollo de la Lógica*, Tecnos, Madrid, 1980.
Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P. y Wang, Y. (1988), «Expertise in a complex skill: Diagnosing x-ray pictures», en M. T. H. Chi, R. Glaser y M. J. Farr (eds.), *The Nature of Expertise*, Lawrence Erlbaum Associates, Hillsdale, N. J.
Lewis, C. I. y Langford, C. H. (1959), *Symbolic Logic*, Dover, Nueva Cork. (Primera edición 1932.)
Lukasiewics, J. (1920), «I. On the Notion of Possibility. II. On Three-Valued Logic», en Storrs McCall (comp.) (1967), *Polish Logic. 1920-1939*, Oxford U.P., Oxford, 40-65. En esa misma compilación, (1930), «Philosophical Remarks on Many-Valued Systems of Propositional Logic», 15-18, y «On Determinism», 19-39.
Marconi, D. (1981), «Types of Non-Scotian Logic»: *Logique et Analyse* (NS), año XXIV, n.º 95-96, sept.-dic., 407-414.
Mendelson, E. (³1987), *Introduction to Mathematical Logic*, Wadsworth & Brooks/Cole.
Morado, R. y Beuchot, M. (1989), «Las paradojas de la implicación material y de la implicación estricta en el siglo XX»: *Analogía*, vol. 3, n.º 1, 63-74.
Morreau, M. (1988), «Default formalisms forgenerics», en M. Krifka (ed.), *Genericity in Natural Language: proceedings of the 1988 Tübingen Conference*. Seminar für natürlichsprachliche Systeme der Universität Tübingen SNS-Bericht 88-42, 337-356.
Nute, D. (1990), «Review of Matthew L. Ginsberg *Readings in Nonmonotonic Reasoning*», 351-355.
Orayen, R. (1995), «Lógica Modal», en *Enciclopedia Iberoamericana de Filosofía*, vol. 7: *Lógica*, Trotta, Madrid, 289-322.

Papalaskari, M.-A. (1988), *Minimal Consequence: A Semantic Approach To Reasoning With Incomplete Information*, University of Edinburgh Ph. D. Dissertation.
Papalaskari, M.-A. y Weinstein, S. (1990), «Minimal consequence in sentential logic»: *Journal of Logic Programming* 9, 19-31. Una versión ligeramente revisada del reporte técnico del mismo título, MS-CIS-87-43, LINC LAB 66, Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, mayo 1987.
Quesada, D. (1995), «Lógica Clásica de Primer Orden», en *Enciclopedia Iberoamericana de Filosofía*, vol. 7: *Lógica*, Trotta, Madrid, 71-104.
Reiter, R. (1978), «On closed world data bases», en Gallaire, Hervé y Minker, J. (eds.), *Logic and Databases*, Plenum Press, New York, 55-76. (Reimpreso en Ginsberg [1987], 300-310.)
Reiter, R. y Criscuolo, G. (1981), «On interacting defaults», en *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, IJCAI-81, 24-28 agosto 1981, University of British Columbia, Vancouver, B. C., Canada. Ann Drinan (ed.), AAAI, Morgan Kaufmann, 270-276. (Reimpreso in Ginsberg [1987], 94-100.)
Robles, J. A. (1995), «Historia de la Lógica». *Enciclopedia Iberoamericana de Filosofía*, vol. 7: *Lógica*, Trotta, Madrid, 1995, 49-69.
Senger, Ch. J. (1989), *Learning Legal Reasoning In Law School: The Differences Between First And Third Year Students*, Ph. D. Dissertation, Michigan State University.
Spade, P. V. (1981), «Three Theories of Obligationes: Burley, Kilvington, and Swyneshed on Counterfactual Reasoning»: *History and Philosophy of Logic* 3, 1-32.
Stein, L. A. (1990), *Resolving Ambiguity in Nonmonotonic Reasoning*, Brown University Ph. D. Dissertation, entregada en mayo de 1991. CS-90-18.
Swyneshed, R. (1979), «Roger Swyneshed's *Insolubilia*: Edition and Comments», editado por Paul Vincent Spade, *Archives d'Histoire Doctrinale et Littéraire du Moyen Âge* 46, 177-220.
Tarski, A. (1930a), «On some fundamental concepts of metamathematics», en Tarski (1983), 30-37.
Tarski, A. (1930b), «Fundamental concepts of the methodology of the deductive sciences», en Tarski (1983), 60-109.
Tarski, A. (1935), «Foundations of the calculus of systems», 342-383 de *Logic, Semantics, Metamathematics: Papers from 1923 to 1938 by Alfred Tarski,* 1983, John Corcoran (ed.), Hackett Publishing Company, segunda edición.
Tarski, A. (1983), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938 by Alfred Tarski*, en John Corcoran (ed.), Hackett Publishing Company, segunda edición.
ter Meulen, A. (1986), «Generic information, conditional contexts and constraints», en Traugott, ter Meulen, Reilly y Ferguson (1986), 123-145.
Troelstra, A. S. (1991), *Lectures in Linear Logic*, CSLI, California.
von Wright, G. H. (1951), *An Essay in Modal Logic*, Amsterdam.

von Wright, G. H. (1963), *Norm and Action*. London.
Wojcicki, R. (1988), «An Axiomatic treatment of non-monotonic arguments»: *Bulletin of the Section of Logic*, vol. 17, n.º 2, 56-61.
Zadeh, L. (1965), «Fuzzy Sets»: *Information and Control* 8, 338-353.

ÍNDICE ANALÍTICO

ÍNDICE DE NOMBRES

NOTA BIOGRÁFICA DE AUTORES

Ignacio Angelelli nació en Roma en 1933. Doctor en Filosofía por la Universidad de Friburgo (Suiza). Es profesor en el Departamento de Filosofía de la Universidad de Texas en Austin. Especialista en Historia y Filosofía de la Lógica y en Filosofía Moderna. Ha publicado ***Studies on Gottlob Frege and traditional philosophy*** (1967), «The techniques of disputation in the history of logic»: *Journal of Philosophy* LXVII (1970), «The substitutivity of identicals in the history of logic», en M. Schirn (ed.), ***Studien zu Frege*** (1976), «Abstraction, looking-around and semantics»: ***Studia Leibnitiana*** (1979) y «Saccheri's postulate»: ***Vivarium*** XXXIII (1995).

Eduardo Alejandro Barrio nació en Buenos Aires en 1965. Especialista en Filosofía de la Lógica y Semántica filosófica, es investigador y profesor del Instituto de Filosofía de la Universidad de Buenos Aires y director de la colección «Enciclopedia lógica». Entre sus obras cabe destacar ***La verdad desestructurada*** (1998), «Verdad y circularidad. El problema de la supervivencia semántica»: ***Theoria*** 43 (2002), «El deflacionismo y el problema de la verdad»: ***Revista Latinoamericana de Filosofía*** 26/2 (2000) y «Conjeturas semánticas: justificar sin confrontar»: ***Análisis filosófico*** XVIII/2 (1998).

Mario Gómez Torrente (Reus, España 1967). Doctor en Filosofía por la Princeton University, su área de especialización es la Filosofía del Lenguaje y la Filosofía de la Lógica. Es investigador en el Instituto de Investigaciones Filosóficas de la UNAM. Ha publicado entre otros trabajos ***Forma y Modalidad. Una introducción al concepto de consecuencia lógica*** (2001), «The Problem of Logical Constants»: ***Bulletin of Symbolic Logic*** 8 (2002), «Logical Truth and Tarskian Logical Truth»: ***Synthese*** 117 (1998/1999) y «Quotation Revisited»: ***Philosophical Studies*** 102 (2001).

Ignacio Jané Palau (Sant Jaume dels Domenys, Tarragona, 1945). Desarrolla su labor docente en el Departamento de Lógica, Historia y Filosofía de la Ciencia de la Universidad de Barcelona. Está especializado en Lógica y Filo-

sofía de la Matemática. Ha publicado *Álgebras de Boole y Lógica* (1989), «A Critical Appraisal of Second-Order Logic»: *History and Philosophyy of Logic* 14 (1993), «The Role of the Absolute Infinite in Cantor's Conception of Set»: *Erkenntnis* 42 (1995), «Reflections on Skolem's Relativity of Set-Theoretical Concepts»: *Philosophia Mathematica* 9 (2001) y «Remarks on Second-Order Consequence»: *Theoria* 18 (2003).

Rodolfo Luján Gaeta (Buenos Aires, 1945). Es especialista en Filosofía de la Ciencia y desarrolla su labor docente en la Facultad de Filosofía y Letras de la Universidad de Buenos Aires. Es autor, entre otras publicaciones, de «Sobre una presunta inconsistencia acerca de la noción de función en la doctrina de Frege»: *Análisis Filosófico* VII/1 (1987), «Significado, referencia e inconmensurabilidad»: *Revista Latinoamericana de* Filosofía XVI/1 (1990), *Lenguaje, identidad y necesidad* (1997), «Realismo y ciencia»: *Episteme* 10 (2000) y «Reasons and Obstacles for a Logic of Discovery»: *Discovery and Creativity* 64/2 (con Nélida Gentile) (1999).

María Gracia Manzano Arjona (Archidona, Málaga, 1950). Es catedrática en el Departamento de Filosofía y Lógica y Filosofía de la Ciencia en la Universidad de Salamanca. Sus principales líneas de investigación son Teoría de los modelos, Lógicas de orden superior y Lógica heterogénea. Es autora, entre otros trabajos, de *Teoría de modelos* (1989), *Extension of First-Order Logic* (1996) y «Life, Work and some miracles of Alonzo Church»: *The Journal of History and Philosophy of Logic* 18 (1997).

Raymundo Morado Estrada (México, D.F., 1959). Doctor en Filosofía por la Universidad de Indiana en Bloomington, es miembro fundador de la Academia Mexicana de Lógica e investigador en el Instituto de Investigaciones Filosóficas de la UNAM en México. Entre sus principales publicaciones se encuentran «La rivalidad en Lógica»: *Dianoia* 30 (1984), «Frege, Hempel and Dedekind: Definition of Number and Correferentiality»: *Ergo* I/2 (1987), «Momentos e Intervalos: Problemas Filosóficos en Lógica Temporal»: *Análisis Filosófico* 18/l (1998) y «Rationality, Logic and Heuristics» (con Leach Savion) (IC-AI'02), Las Vegas (2002).

Alberto Moretti, natural de Avellaneda, Argentina. Especialista en Semántica filosófica y Filosofía de la Lógica. Actualmente es profesor en el Departamento de Filosofía de la Universidad de Buenos Aires y miembro del Consejo Nacional de Investigaciones Científicas. Es autor de *Lo razonable, lo caritativo y el relativismo* (1993), *Fidelidad a los hechos y suspicacia semántica* (1995), «Verdad, paradojas y semántica», en *Filosofía del Lenguaje* I. *Semántica* (EIAF 16) (1995) y *La paradoja de Orayen* (con Guillermo Hurtado) (2003).

Jesús Mosterín Heras (Bilbao, 1941). Es especialista en Lógica, Filosofía de la Ciencia, Teoría de la Información y Teoría de la Racionalidad. Es catedrático en el Departamento de Lógica, Historia y Filosofía de la Ciencia de la Uni-

versidad de Barcelona y profesor de Investigación en el Instituto de Filosofía del CSIC. Autor de *Lógica de Primer Orden* (1983), *Conceptos y Teorías en la Ciencia* (1987), *Filosofía de la Cultura* (1993), *Los lógicos* (2000) y *Diccionario de lógica y filosofía de la ciencia* (con Roberto Torreti) (2002), además de editor de las *Obras completas* de Kurt Gödel (1989).

Raúl Orayen (Avellaneda, Argentina, 1943-México, 2003). Fue profesor de Filosofía en la Universidad de Buenos Aires y posteriormente investigador del Instituto de Investigaciones Filosóficas de la UNAM. Especialista en Lógica, Filosofía de la Lógica y de la Matemática. Además de numerosos artículos publicó *La ontología de Frege* (1971), *Lógica, significado y ontología* (1989), y la colaboración «Lógica modal», en *Lógica* (EIAF, 7) (1995).

Gladys Palau (1937). Es profesora en la Facultad de Filosofía y Letras de la Universidad de Buenos Aires. Sus áreas de especialización son la Lógica y la Filosofía de la Lógica. Entre sus publicaciones destacan *Introducción a la lógica operatoria de Piaget* (1981), *Modelos lógicos y prueba empírica* (con A. Castorina) (1984), *Lógicas divergentes y principios lógicos* (1990), «Psicología y Modelos lógicos», en *Desarrollos en Psicología Genética* (2000) e *Introducción filosófica a las lógicas no clásicas* (2002).

Agustín Rayo Fierro (Ciudad de México, 1973). Especializado en Filosofía de las Matemáticas, Metafísica y Filosofía del Lenguaje. Profesor en la Universidad de St. Andrews en Escocia desde donde regresará a su puesto en la Universidad de California en San Diego. Entre sus publicaciones más importantes figuran «Frege's Correlation»: *Análisis* 64/2 (2004), «Word and Objects»: *Noûs* 36/3 (2002), «Nominalism through D-Nominalization» (con Steve Yablo): *Noûs* 35/1 (2001), «A Puzzle about *de rebus* Belief» (con Vann McGee): *Analysis* 60/4 (2000), «Toward a Theory of Second-order Consequence» (con Gabriel Uzquiano): *The Notre Dame Journal of Formal Logic* 40/3 (1999).

Francisco Rodríguez Consuegra nació en Sevilla en 1951. Profesor de Filosofía en el Departamento de Lógica y Filosofía de la Ciencia de la Universidad de Valencia, es especialista en Filosofía del Lenguaje. De entre sus publicaciones cabe destacar *The Mathematical Philosophy of Bertrand Russell* (1991), y la edición de K. Gödel, *Ensayos inéditos* (1994). Ha traducido a Russell y a Quine al castellano.

El presente volumen pretende ofrecer un panorama representativo de tipos de problemas y orientaciones característicos de la Filosofía de la lógica. Desde la relación de los lenguajes formales con el lenguaje ordinario y los procesos psíquicos, pasando por el estudio de nociones fundamentales como las de identidad, referencia, verdad o consecuencia lógica, hasta las cuestiones de la abstracción y las entidades abstractas, o la consideración de los problemas implicados en la teoría de tipos, la teoría de conjuntos o los distintos sistemas de lógica, los autores hacen una valiosa contribución a la indagación filosófica motivada por la lógica.

Así, el artículo de Agustín Rayo plantea el problema del vínculo entre los razonamientos en un lenguaje constituido social e históricamente y su examen por medio del lenguaje formal. Por su parte, Gladys Palau traza el recorrido histórico de los intentos por fundar la lógica en la investigación psicológica. Rodolfo Gaeta dedica su trabajo a las dificultades de la relación entre identidad y referencia, mientras que Alberto Moretti analiza en el suyo el concepto tarskiano de verdad. La búsqueda de un concepto adecuado de consecuencia lógica es el objeto del artículo de Mario Gómez Torrente. A los enfoques de la abstracción en el siglo XX y a la discusión sobre la existencia de las entidades abstractas consagran sus colaboraciones Ignacio Angelelli y Eduardo Barrio, respectivamente. A partir de las iniciales aportaciones de G. Frege y B. Russell, Francisco Rodríguez Consuegra dedica su contribución a la teoría de tipos, Jesús Mosterín brinda un panorama de la riqueza conceptual de las teorías de conjuntos e Ignacio Jané intenta clarificar qué son esas entidades que denominamos conjuntos. Cierran el volumen los trabajos de María Manzano y de Raymundo Morado, que giran en torno a la aplicabilidad general de la idea de consecuencia lógica y a la tarea de la comparación de sistemas de lógica distintos.

Colaboradores

Alberto Moretti
Raúl Orayen
Agustín Rayo
Gladys Palau
Rodolfo Gaeta
Mario Gómez Torrente
Ignacio Angelelli
Eduardo Barrio
Francisco Rodríguez Consuegra
Jesús Mosterín
Ignacio Jané
María Manzano
Raymundo Morado

EDITORIAL TROTTA



MINISTERIO
DE EDUCACIÓN
Y CIENCIA



CONSEJO SUPERIOR
DE INVESTIGACIONES
CIENTÍFICAS

ISBN 84 - 8164 - 673 - 3

